REMARKS

Claims 30, 31, 32, 44, 45, and 46 have been cancelled. Claims 1-29, 33-43, and 47-53 are now pending in the application. Claims 1, 3, 4, 6, 8, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 25, 27, 33, 34, 35, 36, 37, 40, 41, 42, 43, 47, 48, and 49 have been amended. New claims 50, 51, 52, and 53 have been added. No new matter has been added by amendment. Reexamination and reconsideration of the claims as amended are respectfully requested.

Claim Objections

1.   The Examiner objects to claims 8 and 27 for being in improper form because a multiple dependent claim should refer to other claim in the alternative only. Claims 8 and 27 have been amended and are now in proper form.

Claim Rejections – 35 USC § 112, second paragraph

2.   The Examiner rejects claims 1-49 under 35 U.S.C. 112 second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention.

The Examiner rejects claims 1, 6, 12, 14, 21, 25, 31, 33, 37, 40-42, and 44-46 and the claims that are dependent thereon as indefinite due to the recitation of "PH7JB ". The Examiner states that, "Amending claims 1, 6, 21, 25, 37, and 40 to recite the ATCC deposit number in which seed of corn inbred line PH7JB has been deposited would overcome the rejection." Claims 1, 6, 21, 25, 37, and 40 have been so amended and the ATCC deposit receipt is attached to this response.

The Examiner rejects claims 3 and 22 for the recitation of "wherein the plant is male sterile". The Applicants traverse the rejection but have amended the claims for clarification purposes and to expedite prosecution. Claim 3 has been amended by replacing "male sterile" with --detasseled-- as suggested by the Examiner. The Examiner suggests that a new claim be directed towards a method of producing a male sterile maize plant comprising transforming the maize plant of claim 2 with a nucleic acid that confers male sterility, and another

9

SN: 09/759,800

claim directed toward the male-sterile plant produced by the method of transforming. Claim 22 has been amended and now reads, "The maize plant of claim 2, wherein genes controlling male sterility have been transferred into said maize plant through crossing, wherein PH7JB is a recurrent parent, and wherein said plant has essentially the same morphology and physiology of inbred maize line PH7JB other than the trait of male sterility." Starting on page 1, line 35 and going through line 14 of page 3 of the specification it states that various genes, nuclear and cytoplasmic, have been used to control sterility in maize plants. In the specification on page 4, lines 7-13, it states, "Backcrossing can be used to transfer a specific desirable trait from one inbred or source to an inbred that lacks that trait. This can be accomplished, for example, by first crossing a superior inbred (recurrent parent) to a donor inbred (non-recurrent parent), that carries the appropriate gene(s) for the trait in question. The progeny of this cross is then mated back to the superior recurrent parent followed by selection in the resultant progeny for the desired trait to be transferred from the non-recurrent parent." The technique of backcrossing male sterility genes into an inbred maize plant is well known and well understood to one of ordinary skill in the art. The technique has been successively used since the 1950's (see pages 585-586 of Wych, 1988 included in the Information Disclosure Statement). The amendments contain no new matter. The Applicants request reexamination and reconsideration of the claims as amended.

The Examiner rejects claims 5 and 24 because there is no antecedent basis for "protoplasts". The Applicants have amended claims 4 and 23. The claims now read, "A tissue culture of regenerable cells or protoplasts from the plant of claim 2 {21}." Thus the term "protoplasts" in claims 5 and 24 which depend from claims 4 and 23 respectively, has proper antecedent basis. The amendments place the claims in condition for allowance.

The Examiner rejects claims 14, 33, 41, 45, and 46 because they contain terms such as "high," and "above average" which "do not reasonably apprise one of the scope of the invention." Claims 45 and 46 have been cancelled. Claims 33 and 41 have been amended and no longer include such terms or traits.

SN: 09/759,800

Claim 14 has been amended and no longer includes such terms as "high" and "above average". Claim 14 now reads, "A maize plant, or parts thereof, wherein at least one ancestor of said maize plant is the maize plant of claim 2, said maize plant expressing a combination of at least two PH7JB traits which are not significantly different from PH7JB traits when determined at a 5% significance level and when grown in the same environmental conditions, said PH7JB traits selected from the group consisting of: a relative maturity of 99 based on the Comparative Relative Maturity Rating System for harvest moisture of grain, grain yield, cold test results, and resistance to Northern Leaf Blight." Applicants point out that claim 14 has been amended to clearly define the traits of PH7JB that could be found in a maize plant produced from PH7JB. Applicants have amended the claim using the term "not significantly different from PH7JB traits when determined at a 5% significance level..." as a definitive term. In the specification pages 38-54, the tables show mean trait values. The standards against which the listed traits should be compared are the mean values for those traits exhibited by PH7JB or a maize plant produced from PH7JB in a side-by-side comparison or under other similar environmental conditions. For example, on page 36 lines 17-19 of the specification it discusses that inbred PH7JB produces significantly higher grain yield than inbred PHP02 when both inbreds are crossed to the same tester lines. Applicants also point out that one of ordinary skill in the art of plant breeding would know how to evaluate the traits of two inbred maize lines to determine if they are not significantly different to a 5% significance level in the expression of a given trait. On pages 275-276 in Principles of Cultivar Development (1987) Fehr writes "Two or more independent comparisons of lines in a test provide a means of estimating whether variation in performance among lines is due to differences in genetic potential or to environmental variation." A copy of Fehr, pages 261-286, is attached to this Amendment and Request for Reconsideration as Appendix A. As was done by the Applicants in the specification, mean trait values would be used to determine whether the trait differences are significant. Further, the claim, as amended,

11

SN: 09/759,800

requires that the traits be measured on plants grown in the same environmental conditions.

The Examiner rejects claims 16 and 35 because of improper antecedent basis and has suggested that the recitation "corn plant breeding program" be replaced with --method--. Claims have been amended as suggested by the Examiner thus placing them in condition for allowance.

The Examiner has rejected claims 19, 20, 48 and 49 for improper antecedent basis. Claims 19, 20, 48 and 49 have been amended and are now in proper form for allowance.

The Examiner rejects claim 40 because of the recitation of the word "comprising". The Examiner states that the claim is indefinite because it "does not clearly indicated how many crosses are to be performed by the method." Claim 40 has been amended to specify the first generation (F1) plant. Dependent claim 41 has also been amended to reflect the change in claim 40.

Claim Rejections – 35 USC § 112, first paragraph

3.   The Examiner rejects claims 9-20, 28-39, 41-49 are rejected under 35 U.S.C. 112, first paragraph, as containing subject matter which was not described in the specification in such a way as to reasonably convey to one skilled in the art that the inventors, at the time the application was filed, had possession of the claimed invention. Applicants traverse the rejection.

The Examiner rejects claims 9, 10, 28, and 29 that are directed to F1 hybrids produced with PH7JB as a parent. Applicants note that a claim to the F1 hybrid made with a deposited inbred was expressly acknowledged without reservation by the United States Supreme Court in *J.E.M. Ag. Supply, Inc. v. Pioneer Hi-Bred Int'l, Inc.*, 60 USPQ 2d 1865,1873 (S.Ct. 2001), when the Supreme Court wrote, "...a utility patent on an inbred plant line protects the line as well as all hybrids produced by crossing that inbred with another plant line."

Furthermore, one of ordinary skill in the art would know if they were using or one could easily identify if they were using PH7JB. All F1 plants would have essentially the same genetic markers as the deposited PH7JB. It is well known

12

SN: 09/759,800

to anyone skilled in the art that a hybrid has a genome with one set of the alleles from each inbred. Therefore the genetic profile exhibited in the deposit would be exhibited in the hybrid. As stated in the specification on page 16, lines 8-23, there are many laboratory-based techniques available for the analysis comparison and characterization of plant genotype such as Restriction Length Polymorphisms (RFLPs) and Simple Sequence Repeats (SSRs). Such techniques have been known for some time and may be used to identify whether or not PH7JB was used to develop a hybrid. Applicants also submit to the Examiner the journal article by Berry et al. (2002). This article discusses the probability of identifying the parents of the hybrid by SSR data when neither parent is known. A copy of article by Berry et al. is attached to this Amendment and Request for Reconsideration as Appendix B. The results of the experiment showed that using 100 SSR loci markers resulted in correct parental ranking of inbreds for 53 out of 54 hybrids. Applicants also point out that any breeder of ordinary skill in the art will know the identity of both parents used to produce a hybrid.

The Examiner broadly rejects product claims encompassing any modification of PH7JB, no matter how minor the modification or routine the modification is for a breeder of ordinary skill in the art to make.

As noted in the specification, the development of an inbred line is a time consuming and labor intensive activity. On average, between 10,000 to 20,000 lines are created and screened in order to develop any maize inbred line for which Applicants file a patent application. Once developed, the inbred line is useful for two purposes: (1) to make commercial hybrids, and (2) as a source of breeding material for the development of new inbreds that retain its desired characteristics. A breeder desiring to make a line with similar traits to PH7JB would be greatly advantaged by being able to use PH7JB as starting material. This is because the linked genes arranged through Applicants' breeding efforts, and fixed in PH7JB, can be maintained in the progeny of PH7JB by a breeder of ordinary skill in the art. For example, if a breeder of ordinary skill in the art desired an early maturity version of PH7JB, the breeder could cross PH7JB to an

13

SN: 09/759,800

earlier maturing variety, select for progeny with at least two desired PH7JB traits that also express early maturity, and continue selecting for the traits of PH7JB combined with early maturity. Optionally, the breeder could backcross to PH7JB to obtain further genetic contribution from PH7JB. The end result is the development of an inbred line with substantially all of the benefit of Applicants' work but with only a fraction of the effort.

Specifically, in rejecting the claims for lack of written description, the Examiner states, "the specification also does not describe the plants produced by the corn breeding programs, transgenic PH7JB plants, PH7JB plants comprising single gene conversion(s), or by crosses wherein at least one ancestor is the corn variety PH7JB, other than PH7JB/PH54M. The morphological and physiological traits of the corn plants that are crossed with PH7JB, and with progeny of that cross, are unknown, and the description of progeny and descendents of corn plant PH7JB are unknown. The description of PH7JB is not indicative of the description of plants and seed produced by the breeding programs and crosses, or any of its descendents. The claimed invention also encompasses plants that express at least two of the 'PH7JB traits' listed in claims 14, 33, 41, 45, and 46. However, to say that a plant expresses two traits of another plant is not sufficient information to describe that plant, as numerous corn plants express at least two of the same traits as those expressed by PH7JB. Two plant traits do not provide any description of the other traits of the plant. It is possible that the claimed plants inherited the genes governing those traits from an ancestor other than plant PH7JB. For, example, Piper (U.S. Patent No. 6,188,001) describes a corn plant, designated 'PH1W0,' which has at least two traits in common with PH7JB, high yields, above average cold test results and adapted to the Northcentral region of the United States, for example (col. 10, lines 50-62). The instantly claimed corn plants could have PH1W0 as an ancestor, as well as PH7JB, in which case the dry down and flowering traits, for example, could have been inherited from PH1W0. The claims also encompass plants that do not have to express any of the traits that are expressed by PH7JB."

14

SN: 09/759,800

Applicants note that Examiner's comments represent a change of patent office policy. In numerous previous cases involving the protection of germplasm and progeny claims, including cases allowed after the recently adopted written description guidelines, the listing of traits was previously required by the patent office as a way to meet the written description requirement with respect to progeny. One reason for using traits as a means of description is because it was, and still is, technologically impossible to sequence the entire genome of a specific variety.

If it was possible to sequence the genome of a variety, PH7JB could be described and compared to the prior art to identify its unique genetic sequences and sequence combinations, and presumably, claims to progeny retaining those unique genetic aspects would be allowed by the patent office. This would be analagous to the way claims are examined for individual short genetic sequences and claims allowed for any plant comprising a specific transgene. Applicants assert that the fact that technological tools do not exist to fully describe the unique characteristics of the full genome of PH7JB does not make the progeny lines derived therefrom any less entitled to adequate patent protection. It is the purpose of the patent law to protect new and useful processes, compositions of matter and improvements thereof. 35USC 101.

This situation is somewhat analagous to Ex Parte Tanksley, 37 USPQ2d. 1382. In that case the Examiner desired that Tanksley claim according to sequence data to "better characterize the cDNA clones" and "facilitate a complete search of the prior art" and issued a 112 first paragraph written description rejection. The Board held that "the section 112 rejection amounts to a requirement...that the appellants amend their claims in a specified manner...We find no language in the statute or case law which would support that requirement." The Board, in treating the section 112 first paragraph rejection as a 112 second paragraph rejection, held that "In our judgement, a patent applicant is entitled to a reasonable degree of latitude in complying with the second paragraph of 35 U.S.C. 112 and the examiner may not dictate the literal terms of the claims for the stated purpose of facilitating a search of the

15

SN: 09/759,800

prior art. Stated another way, a patent applicant must comply with 35 U.S.C. 112, second paragraph, but just how the applicant does so, within reason, is within applicant's discretion." Id. at 1386.

Applicants have amended claims 17, 33 and 36 to limit the progeny covered to those within a pedigree distance of two crosses away from PH7JB. Claim 41 is limited to one cross away from PH7JB by virtue of dependency. Within the plant breeding arts breeders use pedigree as a means to characterize lines in reference to their progenitors. To those of ordinary skill in the art, this indicates that a line fewer crosses away from a starting line will be, as a whole, more highly related to the starting line. Thus, the work of the original breeder in developing the starting line will be retained in the closely related progeny. More specifically, traits and linkage groups present in PH7JB will be retained in progeny that are within 2 outcrosses from PH7JB. Applicants submit that characterization of the progeny of PH7JB by virtue of their filial relationship is clearly within reason. Not only are filial descriptions used by breeders to evaluate materials for use in their breeding programs, but it is standard practice within the plant breeding industry for licensor's of inbred maize lines to retain a royalty from lines developed through the use of their inbreds. Those royalties are, in almost all cases, based on the filial relationship between the licensed inbred used in breeding and the progeny line commercialized. This provides evidence that those of ordinary skill in the art of plant breeding describe progeny in terms of pedigree.

Applicants also note that the mere fact that the progeny have not been created does not prevent them from being patented. As stated in MPEP 2163 (3) (a), "An invention may be complete and ready for patenting before it has actually been reduced to practice." As stated in the written description guidelines "an applicant shows possession of the claimed invention by describing the claimed invention with all its limitations using such descriptive means as words, structures, figures, diagrams, and formulas that fully set forth the claimed invention. Possession may be shown in a variety of ways, including...by describing distinguishing identifying characteristics sufficient to show that the

16

SN: 09/759,800

applicant was in possession of the claimed invention." 1255 Official Gazette 140 (Feb. 5, 2002). P digre , which is a formula used by plant breeders, is a distinguishing identifying characteristic in compliance with the written description guidelines. Further, the Examiner must evaluate written description by the claimed invention with all of its limitations, including the limitation of being derived from PH7JB.

PH7JB-derived progeny are described by the fact that PH7JB is utilized in a breeding program to make the PH7JB-derived progeny, PH7JB gives genetic contribution to the PH7JB-derived progeny, and the genetics of PH7JB are described by ATCC deposit of PH7JB seed. By limiting the progeny to 2 or less crosses away from PH7JB, the Examiner's concern that the progeny may be only distantly related to PH7JB is addressed. In Enzo vs. Gen-Probe, U.S. State Court of Appeals for the Federal Circuit, 63 USPQ 2d 1609, the court reversed its prior decision regarding the insufficiency of the deposited genetic probes to meet the written description requirement. In so holding, the court stated, " As the deposited sequences are about 850, 8500, and 1300 nucleotides long, ..., there are at least hundreds of subsequences of the deposited sequences, an unknown number of which might also meet the claimed hybridization ratio. Moreover, Enzo's expert, Dr. Wetmur, stated that 'astronomical' numbers of mutated variations of the deposited sequence also fall within the scope of those claims, and that such broad claim scope is necessary to adequately protect Enzo's invention from copyists who could otherwise make minor change to the sequence and thereby avoid infringement while still exploiting the benefits of Enzo's invention. The defendants assert that such breadth is fatal to the adequacy of the written description. On the other hand, because the deposited sequences are described by virtue of a reference to their having been deposited, it may well be that various subsequences, mutations, and mixtures of those sequences are also described to one of skill in the art. We regard that question as an issue of fact...."

The issue of whether the progeny as now claimed satisfies the written description requirement is also an issue of fact. One of ordinary skill in the art

17

would know if PH7JB were utilized in a breeding program by looking at the breeding records and therefore would know if a progeny were derived from PH7JB. PH7JB is a unique Inbred, as evidenced by the morphological and physiological traits given in Table 1, pages 18-20, of the application. Routinely used molecular techniques, discussed on page 16, lines 8-23 of the application, can be used to verify whether PH7JB is within the pedigree of a line.

Applicants would also like to emphasize that PH7JB cannot be derived through any other means then through PH7JB seed and plant, nor can the influence of PH7JB on the progeny be removed from a line within 2 outcrosses of PH7JB. This fact also highlights the different perspective between the Examiner and the Applicants regarding the scope of the claims. The Examiner believes the claims to progeny to be of great breadth. However, to view these claims as being of great breadth merely because a large number of plants could theoretically fall within its scope ignores an essential limitation of the claim; that only a plant developed through the use of PH7JB is within the scope of the claim. Such a plant could not be independently derived without the use of PH7JB, so the claim would not in any way restrict the work of a breeder that did not in fact use PH7JB. A breeder infringing such a claim must have made a conscious choice to use PH7JB in order to obtain some or all of PH7JB's desired characteristics. Compliance with the written description requirement is essentially a fact-based inquiry that will "necessarily vary depending on the nature of the invention claimed." Vas-Cath v. Mahurkar, 935 F. 2d 1555 (citing In re DiLeone, 436 F2d. 1404, 1405). Thus, the compliance with the written description requirement must be judged in view of this limited scope of the progeny claims. As amended, the claims are drawn to only a limited scope of progeny, progeny which but for Applicants' creation of PH7JB could never have existed. This is in harmony with the statement in section 2163 of the MPEP that "the written description requirement promotes the progress of the useful arts by ensuring inventions are adequately described in the specification in exchange for the right to exclude." That quid pro quo of patent law has been met by the Applicants in the present case, and to use written description to deny adequate

18

SN: 09/759,800

patent protection would be contrary to the stated purpose of the written description requirement.

The Examiner goes on to reject claims to PH7JB plants further containing transgenes and single gene conversions under 35 U.S.C.112, first paragraph. Applicants note that examples of traits and single gene conversions are given in the specification, page 22, lines 33 through 35 line 2. Even if more than one trait is affected by the transgene, the genetics of PH7JB will be only minimally affected. The Examiner must consider all limitations of the claimed invention. While the Examiner is focusing on traits, the Applicants point out that they are not claiming so broadly as to claim any maize plant, regardless of source, comprising those traits. Applicants are claiming PH7JB, or a limited set of plants derived therefrom, that retain significant features of PH7JB. Applicants have made an enabling deposit of PH7JB with the ATCC, and Applicants are seeking a fair scope of protection as the quid pro quo for the teaching in the specification and the deposit of the material. The insertion of one or a few genes into a genome that is estimated to have over 50,000 to 80,000 genes (Xiaowu, Gai et al., Nucleic Acids Research, 2000, Vol. 28, No. 1, 94-96) is a minor change to PH7JB and will not prevent one of skill in the art from identifying the plant as PH7JB. In addition, to expedite prosecution, Applicants have amended claim 11. Claim 11 now reads, "The maize plant, or parts thereof, of claim 2, wherein the plant or parts thereof have been transformed so that its genetic material contains one or more transgenes that confer a qualitative trait." Qualitative traits, as described in an introductory plant breeding book, are traits that "have phenotypes that can be divided into discrete classes...They are controlled by one or a few major genes whose expression is not influenced markedly by the environment" (Fehr, W., Principles of Cultivar Development, vol.1, 1987, page 26). Claim 30 has been cancelled. The Examiner has suggested that claims 11 and 30 be amended to list the types of transgenes contemplated in the specification, for example disease or pest resistance genes, provided the prior art teaches those isolated genes. The Applicants believe that an amendment as suggested by the Examiner is limiting the scope to which the Applicants are

19

SN: 09/759,800

entitled. Examples of specific transgenes are given in the specification and one of ordinary skill in the art would be able to determine additional transgenes that may be used. The Examiner states that, "Transgenes may also be of any gene, including those that effect more than one trait. For example, a transgene that is a transcription factor can effect more than just one gene, and multiple traits. Such plants would express different morphological and physiological traits from PH7JB, which are not described." Applicants point out that the molecular profile of such a plant would be substantially unchanged and therefore one would be able to identify such a plant. Applicants have amended claim in order to expedite allowance.

The Examiner rejects claims 12, 13, 31, and 32. Claims 12 and 31 are drawn to the method of crossing a PH7JB plant containing a transgene with another plant. Claims 13 and 32 are to the plant made from the method. Claims 31 and 32 have been cancelled. Claims 12 and 13 have been amended for clarification purposes. Applicants point out that the methods are fully described. Furthermore, one of ordinary skill in the art would know if they were using or one could easily identify if they were using PH7JB or PH7JB further containing a transgene to develop a hybrid. All F1 plants would comprise essentially the same genetic markers as the deposited PH7JB. It is well known to anyone skilled in the art that a hybrid has a genome with one set of the alleles from each inbred. Therefore the genetic profile exhibited in the deposit would be exhibited in the hybrid. The plant of claim 13 would have the genetic profile of PH7JB except at the site of integration of the transgene. The change of one to a few genes out of an estimated 50,000 to 80,000 genes is a minor change and will not prevent one of ordinary skill in the art from identifying the plant as PH7JB. One of ordinary skill in the art would also know how to cross PH7JB containing a transgene with another plant to produce a hybrid. Thus, the Applicants have described the invention with sufficient specificity to enable others to make and use the invention. In light of the arguments and amendments, the Applicants request that the Examiner withdraw his rejection to claims 12 and 13.

20

SN: 09/759,800

The Examiner also rejects claims 37-39 under 35 USC § 112, first paragraph, as containing subject matter which was not described in the specification in such a way as to reasonably convey to one skilled in the relevant art that the inventor, at the time of the application was filed, had possession of the claimed invention. Claims 37-39 are clearly directed to growing out an F(1) hybrid in which PH7JB is a parent and searching for PH7JB inbred seed. Due to the imperfect process of seed production parent seed can sometimes be contained in the hybrid seed bag. The claims merely claim the method of searching for inbred PH7JB seed within a bag of hybrid seed. The method is also clearly described in the specification on page 5, line 21 through line 7 on page 6. The Applicants request that the Examiner withdraw his rejection to claims 37-39.

Lastly, The Examiner has rejected certain method claims under written description. Applicants point out that the methods are fully described, as is the starting material in the method, PH7JB. One of ordinary skill in the art would know how to cross PH7JB to develop an F1 hybrid and also how to self plants derived from crosses with PH7JB for the purpose of developing an inbred plant. In *Ex parte Parks*, 30 USPQ 2d 1234 (B.P.A.I. 1994), the Board of Appeals stated, "Adequate description under the first paragraph of 35 U.S.C. 112 does not require *literal* support for the claimed invention. Rather, it is sufficient if the originally-filed disclosure would have conveyed to one having ordinary skill in the art that an appellant had possession of the concept of what is claimed." Emphasis added. In *J.E.M. Ag. Supply*, the Supreme Court also acknowledged the value of a newly developed line in further breeding, when it stated that, "...a breeder can use a plant that is protected by PVP certificate to 'develop' a new inbred line while he cannot use a plant patented under §101 for such a purpose." Id. at 1873. In light of the amendments to the claims and the foregoing arguments the Applicants request reconsideration of the rejection under the first paragraph of 35 U.S.C. 112.

21

SN: 09/759,800

4.   The Examiner rejects claims 18-20 and 47-49 under 35 U.S.C. 112, first paragraph, as containing subject matter which was not described in the specification in such a way as to enable one skilled in the art to which it pertains, or with which it is most nearly connected, to make and/or use the invention. Applicants traverse the rejection.

The Examiner rejects claims 18-20 and 47-49 that are directed to PH7JB inbred maize plant further comprising one or more single gene conversions. Claims 18 and 47 are directed to PH7JB that contains a gene that has been transferred to PH7JB through traditional breeding methods. The claims have been amended to expedite prosecution. Claim 18 now reads, "The maize plant, or parts thereof, of claim 2, further comprising one or more genes that confer a qualitative trait and have been transferred into said maize plant through breeding methods that utilize PH7JB as a recurrent parent." Claim 47 now reads, "The maize plant, or parts thereof, of claim 2, further comprising one or more genes that have been transferred into said maize plant by utilizing PH7JB as a recurrent parent and wherein the maize plant, or parts thereof, are essentially unchanged from inbred maize line PH7JB." Once again the Applicants would like to point out that one of ordinary skill in the art would be able to detect a PH7JB maize plant that contains genes that have been inserted through crossing. The genetics would be substantially the same as PH7JB as would the morphological and physiological traits of PH7JB. The specification states, "A further embodiment of the invention is a single gene conversion or introgression of the maize plant disclosed herein in which the gene or genes of interest (encoding the desired trait) are introduced through traditional (non-transformation) breeding techniques, such as backcrossing (Hallauer et al, 1988)."

The Examiner has cited articles and states that they "teach that it is unpredictable whether the gene or genes responsible for conferring a phenotype in one plant genotypic background may be introgressed into the genetic background of a different plant, to confer a desired phenotype in said different plant." The Examiner states that, "Hunsperger et al. teach that the introgession

22

SN: 09/759,800

of a gene in one genetic background on any plant of the same species, as performed by sexual hybridization, is unpredictable in producing a single gene conversion plant with a desired trait (column 3, lines 26-46). " This is not what is taught by Hunsperger et al. Hunsperger et al. teaches that a gene that results in dwarfism of a petunia plant can be incorporated into other genetic backgrounds of the petunia species (See column 2, line 67 to column 3, lines 1-4). Hunsperger et al. merely discusses the level of the expression of that gene differed in petunia plants of different genetic backgrounds. Hunsperger et al. succeeded in incorporating the gene into petunia plants of different genetic backgrounds. Therefore Hunsperger et al. demonstrate that one of ordinary skill in the art can use traditional breeding methods to obtain maize plants containing genes that confer a qualitative trait. The specification provides ample disclosure of starting materials such as maize inbred PH7JB, a discussion of traditional breeding methods that may be used, and examples of transgenes and naturally occurring genes. Please note in Hallauer et al. (1988) on page 472, submitted in the information disclosure statement, it states that, "For single gene traits that are relatively easy to classify, the backcross method is effective and relatively easy to manage."

The Examiner goes on to state that, "Kraft et al. teach that linkage disequilibrium effects and linkage drag prevent the making of plants comprising a single gene conversion, and that such effects are unpredictably genotypic specific and loci-dependent in nature (page 323, column 1, lines 7-15)." Applicants disagree that the article states such points. Applicants assume that the Examiner is trying to point out that one gene cannot be introduced into a plant using traditional breeding techniques such as backcrossing without also introducing closely linked genes into the plant. It is well understood in the relevant art that DNA surrounding the gene of the desired trait is introduced into the plant when traditional breeding techniques are utilized to insert a gene into a plant of interest. It is also understood in the art that introducing a gene into a plant variety such as PH7JB is an insubstantial change to the variety. The World Seed Organization, on it's web site, writes, "The concept of an essentially

23

derived variety was introduced into the 1991 Act of the UPOV Convention in order to avoid plagiarism through mutation, multiple back-crossing and to fill the gap between Plant Breeder's Rights and patents." As determined by the UPOV Convention, essentially derived varieties may be obtained for example by the selection of a natural or induced mutant, or of a somaclonal variant, the selection of a variant individual from plants of the initial variety, backcrossing, or transformation by genetic engineering. The commercialization of an essentially derived variety needs the authorization of the owner on the rights vested in the initial variety." International Convention for the Protection of New Varieties of Plants, as amended on March 19, 1991, Chapter V, Article 14, Section 5(c), (emphasis added). A copy of the relevant portion of the UPOV Convention and the World Seed Organization web site is attached as Appendix C.

An example of how one of ordinary skill in the art can transfer a gene conferring a qualitative trait into a variety through backcrossing is demonstrated by the fact that the commercial market now distributes a multitude of products produced in this manner. Such conversion lines are easily developed without undue experimentation. Poehlman et al. (1995) on page 334, submitted in the information disclosure statement, states that, "A backcross-derived inbred line fits into the same hybrid combination as the recurrent parent inbred line and contributes the effect of the additional gene added through the backcross."

The Examiner goes on to state that, "Eshed et al. teach that in plants, epistatic genetic interactions from the various genetic components comprising contributions from different genomes may effect quantitative traits in genetically complex and less than additive fashion (page 1815, column 1, line 1 to page 1816, column 1, line 1). Applicants would like to first point out on page 1816, column 1, lines 1-5 of the Eshed et al. article it states, "Recent studies that detected epistasis of selected QTL in Drosophila (Long et al. 1995), soybean (Lark et al.1995) and maize (Doebley et al.1995; Cockerham and Zeng 1996) did not show a less-than-additive trend." Emphasis added. Applicants also add that transferring a qualitative trait does not require undue experimentation. Please note Hallauer et al. (1988) on page 472, submitted in the information disclosure

24

SN: 09/759,800

statement, which states, "For single gene traits that are relatively easy to classify, the backcross method is effective and relatively easy to manage." As stated previously claims 18 and 47 have been amended to expedite prosecution. In claim 18, the genes transferred into PH7JB are now limited to qualitative traits. Claim 47 is now limited to plants that are essentially unchanged from PH7JB. Given the arguments and the amendments the Applicants request reexamination and reconsideration of the claims.

As noted in the specification, the development of an inbred line is a time consuming and labor intensive activity. On average, between 10,000 to 20,000 lines are created and screened in order to develop any maize inbred line for which the Applicants file a patent application. Once developed, the inbred line is useful for two purposes: (1) to make commercial hybrids, and (2) as a source of breeding material for the development of new inbreds that retain the original inbred's desired characteristics. A breeder desiring to make a line with similar traits to PH7JB would be greatly advantaged by being able to use PH7JB as starting material. This is because the linked genes arranged through Applicants' breeding efforts, and fixed in PH7JB, can be maintained in the progeny of PH7JB by a breeder of ordinary skill in the art. For example, if a breeder of ordinary skill in the art desired a waxy-kernel corn version of PH7JB, the breeder could cross PH7JB to a waxy-kernel corn variety, select for progeny with the desirable traits of PH7JB that also express the waxy kernel trait, and continue selecting for the traits of PH7JB combined with waxy kernels. Optionally, the breeder could backcross to PH7JB to obtain further genetic contribution from PH7JB. The end result is the development of an inbred line with substantially all of the benefit of Applicants' work but with only a fraction of the effort.

5.   The Examiner rejects claims 1-49 under 35 U.S.C. 112, first paragraph, as containing subject matter which was not described in the specification in such a way as to enable one skilled in the art to which it pertains, or with which it is most nearly connected, to make and/or use the invention. The Examiner states that the seeds must be readily available to the public. The Applicants have deposited

25

SN: 09/759,800

the seed of PH7JB with the ATCC and have amended the claims accordingly. A receipt of the deposit is attached. The deposit and amendment to the claims obviate the rejection.

In light of the amendments to the claims and the foregoing arguments the Applicants request reconsideration of the rejection under the first paragraph of 35 U.S.C. 112.

Claim Rejections under 35 U.S.C. § 102 and 103

6. Examiner states that, "Claims 1-49 are rejected under 35 U.S.C. 102(b) as anticipated by or in the alternative, under 35 U.S.C. 103(a) as obvious over Piper (U.S. Patent No. 6,188,001). " Applicants traverse the rejection.

The Examiner goes on to state, "Piper teaches seed of an inbred maize line designated 'PH1W0', plants produced by growing said seed, and plants and plant parts having all the physiological and morphological characteristics of PH1W0 (col. 10, lines 50-62; Table 1; claims). It appears that the claimed plants and seeds of the instant invention may be the same as PH1W0, given that they exhibit the similar traits, high grain yield, above average cold test results, and adapted to the Northcentral region of the United States, for example (col. 10, lines 50-62). Alternatively, if the claimed plants, plant parts, and seeds of PH7JB are not identical to PH1W0, then it appears that PH1W0 only differs from the instantly claimed plants, plant parts, and seeds due to minor morphological variation, wherein said minor morphological variation would be expected to occur in different progeny of the same cultivar, and wherein said minor morphological variation would not confer patentable distinction to PH7JB."

Applicants point out that the designation "PH7JB" of the instantly claimed cultivar is not arbitrarily assigned. PH7JB seed has been deposited with the ATCC and the specification and the appropriate claims have been amended to include the ATCC deposit number. Applicants also point out that the differences between PH7JB and PH1W0 are not "minor morphological variations". On page 17, lines 3-12 of the specification it states that PH7JB is a maize inbred that is best suited to be used as a female in hybrid production. In column 10, line 51-62

26

SN: 09/759,800

of th Piper patent it states that PH1W0 is a maize inbred that is best suited to be us_d as a male or a female. Other diff rences are taken from Table 1 of the specification, pages 18-20, and Table 1 of the Piper patent, columns 11-16 and are listed in the following table.

PH7JB	PH1W0
Anther color is yellow	Anther color is pink
Glume color is purple	Glume color is light green
Silk color is pink	Silk color is red
2 = Anthracnose Stalk rot resistance score	7 = Anthracnose Stalk rot resistance score
5 = Fusarium ear and kernel rot score	6 = Fusarium ear and kernel rot score
5,827 Kg/ha yield	5,284 Kg/ha yield

The examples and the list are not exhaustive but they give ample evidence that the inventions are not the same. Nor are they minor variations of each other.

Applicants have canceled claims 44, 45 and 46. Applicants have amended claim 14. Applicants have amended claims 40, 41, 42, and 43. Claim 41 has been amended and now reads, "A first generation (F1) PH7JB-progeny maize plant, or parts thereof, produced by the method of claim 40." Claim 41 is now one cross away from PH7JB. Claim 41 clearly states that PH7JB must be used to obtain a PH7JB-progeny maize plant. Claim 42 has been amended so that it does not allow any further crosses away from PH7JB. Thus claim 42 is the selfing of the plant derived by the one-cross-away from PH7JB made in claim 40. Claim 43 has been amended for clarification purposes. These PH7JB-progeny plants are limited to one cross away from PH7JB and the progeny plants are limited by the use of PH7JB in the initial cross. Applicants contend that progeny of PH7JB could not the same as PH1W0 or the progeny of PH1W0 because PH7JB is not PH1W0. One would not be able to obtain plants within one cross of PH7JB through modification of the maize inbred PH1W0 taught in U.S. Patent No. 6,188,001 because PH7JB comprises a unique and nonobvious combination of genetics. As evidenced by past arguments and the declaration of Stephen Smith submitted as Appendix D, the claimed progeny plants of PH7JB retain unique and nonobvious combinations of genetics derived from PH7JB. Thus, they deserve to be considered new compositions in their own right.

27

SN: 09/759,800

In light of the above, Applicants respectfully request that the Examiner reconsider and withdraw the rejection to claims 1-49 under 35 U.S.C. 102 (b) and 103(a).

Cancellation of claims 30, 31, 32, 44, 45, and 46; amendment of claims 1, 3, 4, 6, 8, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 25, 27, 33, 34, 35, 36, 37, 40, 41, 42, 43, 47, 48, and 49; and addition of claims 50, 51, 52, and 53 does not in any way change the claim scope which the Applicants believe is allowable but is meant to hasten the issuance of the patent.

CONCLUSION

Attached hereto is a marked-up version of the changes made to the specification and claims by current amendment. The attached page is captioned **"VERSION WITH MARKINGS TO SHOW CHANGES MADE".**

Applicants submit that in light of the foregoing amendments and the remarks, the claims 1-29, 33-43, and 47-53 are in condition for allowance. Reconsideration and early notice of allowability is respectfully requested. If it is felt that it would aid in prosecution, the Examiner is invited to contact the undersigned at the number indicated to discuss any outstanding issues.

Respectfully submitted,

Steven Callistein
Reg. No. 43,525
Attorney for Applicants

Steven Callistein
Pioneer Hi-Bred International
7100 NW 62nd Avenue
P.O. Box 1000
Johnston, IA 50131-1000
(515)-254-2823

28

SN: 09/759,800

VERSION WITH MARKINGS TO SHOW CHANGES MADE

In the Specifications

At page 55, following "<u>Deposits</u>", the entire paragraph, lines 2-21, were deleted and the clean paragraph as written was inserted.

In the Claims

Claims 30, 31, 32, 44, 45 and 46 were cancelled.

Claims 1, 3, 4, 6, 8, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 25, 27, 33, 34, 35, 36, 37, 40, 41, 42, 43, 47, 48 and 49 were amended as follows.

1. (Amended) Seed of maize inbred line designated PH7JB, representative seed of said line having been deposited under ATCC Accession No.[_____] <u>PTA-4531</u>.

3. (Amended) The maize plant of claim 2, wherein said plant is [male sterile] <u>detasseled</u>.

4. (Amended) A tissue culture of regenerable cells-<u>or-protoplasts from the plant</u> of claim 2.

6. (Amended) A maize plant regenerated from the tissue culture of claim 4, capable of expressing all the morphological and physiological characteristics of inbred line PH7JB, representative seed of which have been deposited under ATCC Accession No. [_____] <u>PTA-4531</u>.

8. (Amended) The method of claim 7 wherein [the] inbred maize <u>line PH7JB</u> [plant of claim 2] is the female or male parent.

29

SN: 09/759,800

11. (Amended) The maize plant, or parts thereof, of claim 2, wherein the plant or parts thereof have been transformed so that its genetic material contains one or more transgenes [operably linked to one or more regulatory elements] that confer a qualitative trait.

12. (Amended) A method for producing a first generation (F1) maize plant [that contains in its genetic material one or more transgenes,] comprising crossing the maize plant of claim 11 with [either] a second plant [of another maize line, or a non-transformed maize plant of the line PH7JB, so that the genetic material of the progeny that result from the cross contains the transgene(s) operably linked to a regulatory element].

13. (Amended) [Maize plants] The first generation (F1) maize plant, or parts thereof, produced by the method of claim 12.

14. (Amended) A maize plant, or parts thereof, wherein at least one ancestor of said maize plant is the maize plant of claim 2, said maize plant expressing a combination of at least two PH7JB traits which are not significantly different from PH7JB traits when determined at a 5% significance level and when grown in the same environmental conditions, said PH7JB traits selected from the group consisting of: a relative maturity of [approximately] 99 based on the Comparative Relative Maturity Rating System for harvest moisture of grain, [high] grain yield, [above average] cold test results, and [above average] resistance to Northern Leaf Blight[, and adapted to the Northcentral region of the United States].

15. (Amended) A method for developing a PH7JB-progeny maize plant in a maize plant breeding program using plant breeding techniques, which include employing a maize plant, or its parts, as a source of plant breeding material, comprising: obtaining the maize plant, or its parts, of claim 2 as a source of said breeding material.

30

SN: 09/759,800

16. (Amended) The [maize plant breeding program] method of claim 15 wherein plant breeding techniques are selected from the group consisting of: recurrent selection, backcrossing, pedigree breeding, restriction fragment length polymorphism enhanced selection, genetic marker enhanced selection, and transformation.

17. (Amended) [A] The PH7JB-progeny maize plant, or parts thereof, produced by the method of claim 15 wherein the method comprises 2 or less crosses to a plant other than PH7JB or a plant that has PH7JB as a parent.

18. (Amended) The maize [plants] plant, or parts thereof, of claim 2, further comprising one or more [single gene conversions] genes that confer a qualitative trait and have been transferred into said maize plant through breeding methods that utilize PH7JB as a recurrent parent.

19. (Amended) The [single gene conversion(s)] maze plant of claim 18, wherein [the gene] at least one of the genes is a dominant allele.

20. (Amended) The [single gene conversion(s)] maize plant of claim 18, wherein [the gene] at least one of the genes is a recessive allele.

21. (Amended) A maize plant, or parts thereof, having all the physiological and morphological characteristics of inbred line PH7JB, representative seed of said line having been deposited under ATCC accession No. [_____] PTA-4531.

22. (Amended) The maize plant of claim [21] 2, wherein [said plant is male sterile] genes controlling male sterile have been transferred into said maize plant through crossing, wherein PH7JB is a recurrent parent, and wherein plant has essentially the same morphology and physiology of inbred maize line PH7JB other than the trait of male sterility.

31

23. (Amended) A tissue culture of regenerable cells or protoplasts from the plant of claim 21.

25. (Amended) A maize plant regenerated from the tissue culture of claim 23, capable of expressing all the morphological and physiological characteristics of inbred line PH7JB, representative seed of which have been deposited under ATCC Accession No. [_____] PTA-4531.

27. (Amended) The method of claim 26 wherein the inbred maize plant [of claim 21] having all the morphological and physiological characteristics of inbred maize plant PH7JB is the female or male parent.

33. (Amended) A PH7JB-progeny maize plant, or parts thereof, wherein at least one ancestor of said PH7JB-progeny maize plant is the maize plant of claim [21] 2, [said maize plant expressing a combination of at least two PH7JB traits selected from the group consisting of: a relative maturity of approximately 99 based on the Comparative Relative Maturity Rating System for harvest moisture of grain, high grain yield, above average cold test results, above average resistance to Northern Leaf Blight, and adapted to the Northcentral region of the United States] and wherein the pedigree of said PH7JB-progeny maize plant is within 2 or less crosses to a plant other than PH7JB or a plant that has PH7JB as a parent.

34. (Amended) A method for developing a PH7JB-progeny maize plant in a maize plant breeding program using plant breeding techniques, which include employing a maize plant, or its parts, as a source of plant breeding material, comprising: obtaining the maize plant, or its parts, of claim 21 as a source of said breeding material.

35. (Amended) The [maize plant breeding program] method of claim 34 wherein plant breeding techniques are selected from the group consisting of: recurrent

selection, backcrossing, pedigree breeding, restriction fragment length polymorphism enhanced selection, genetic marker enhanced selection, and transformation.

36.   (Amended) [A] The PH7JB-progeny maize plant, or parts thereof, produced by the method of claim 34 wherein the method comprises 2 or less crosses to a plant other than PH7JB or a plant that has PH7JB as a parent.

37.   (Amended) A process for producing inbred PH7JB, representative seed of which have been deposited under ATCC Accession No. [_____] PTA-4531, comprising:

    (a)   planting a collection of seed comprising seed of a hybrid, one of whose parents is inbred PH7JB said collection also comprising seed of said inbred;

    (b)   growing plants from said collection of seed;

    (c)   identifying said inbred PH7JB plants;

    (d)   selecting said inbred PH7JB plant; and

    (e)   controlling pollination in a manner which preserves the homozygosity of said inbred PH7JB plant.

40.   (Amended) A method for producing a first-generation (F1) PH7JB-[derived] progeny maize plant, comprising:

    (a)   crossing inbred maize line PH7JB, representative seed of said line having been deposited under ATCC Accession No. [_____] PTA-4531, with a second maize plant to yield progeny maize seed;

    (b)   growing said progeny maize seed, under plant growth conditions, to yield said first generation (F1) PH7JB-[derived] progeny maize plant.

41.   (Amended) A first generation (F1) PH7JB-[derived] progeny maize plant, or parts thereof, produced by the method of claim 40[, said PH7JB-derived maize

SN: 09/759,800

plant expressing a combination of at least two PH7JB traits selected from the group consisting of : a relative maturity of approximately 99 based on the Comparative Relative Maturity Rating System for harvest moisture of grain, high grain yield, above average cold test results, above average resistance to Northern Leaf Blight, and adapted to the Northcentral region of the United States].

42. (Amended) [The] A method for producing a PH7JB progeny inbred maize plant, comprising generating the first generation (F1) PH7JB-progeny maize plant by the method of claim 40[,] and further comprising:

[(c) crossing said PH7JB-derived maize plant with itself or another maize plant to yield additional PH7JB-derived progeny maize seed;

(d) growing said progeny maize seed of step (c) under plant growth conditions, to yield additional PH7JB-derived maize plants;

(f) repeating the crossing and growing steps of (c) and (d) from 0 to 5 times to generate further PH7JB-derived maize plants] selfing said first generation (F1) PH7JB-progeny maize plant for successive filial generations to generate a PH7JB inbred progeny maize plant.

43. (Amended) [A further derived] The PH7JB inbred progeny maize plant, or parts thereof, produced by the method of claim 42.

47. (Amended) The maize [plants] plant, or parts thereof, of claim [21] 2, further comprising one or more [single gene conversions] genes that have been transferred into said maize plant by utilizing PH7JB as a recurrent parent and wherein the maize plant, or parts thereof, are essentially unchanged from inbred maize line PH7JB.

48. (Amended) The [single gene conversion(s)] maize plant of claim 47, wherein [the gene] at least one of the genes is a dominant allele.

34

SN: 09/759,800

49.   (Amended) The [single gene conv rsion(s)] <u>maize plant</u> of claim 47,
wherein [the gene] <u>a least one of the genes</u> is a recessive allele.

Claims 50, 51, 52, and 53 were added.

APPENDIX A

PRINCIPLES OF CULTIVAR DEVELOPMENT

VOLUME 1

Theory and Technique

Walter R. Fehr

Iowa State University

with the assistance of
Elinor L. Fehr and Holly J. Jessen

Macmillan Publishing Company
A Division of Macmillan, Inc.
New York

Collier Macmillan Publishers
London

CHAPTER NINETEEN

Field-Plot Techniques

The fundamental purpose of plant breeding is to identify genotypes with superior performance in commercial production. A large proportion of the time and expense devoted to cultivar development is in field evaluation of breeding material. The tests may involve genotypes in an initial stage of evaluation or those being given final consideration for release as new cultivars. The characters evaluated range from those that can be measured readily by visual examination to those that must be measured with appropriate instruments. The genetic potential of a genotype for some characters may be determined effectively with one or a few plants in a small plot, while for other characters extensive evaluation in larger plots may be needed.

It is the responsibility of the plant breeder to select the field-plot techniques that will provide the maximum amount of information with the resources available. The challenge is to adequately test as many genotypes as possible. The resources available to plant breeders vary: usually several alternative techniques are available for character evaluation. Plant breeders must decide which techniques will be the most effective and efficient in their particular situation.

Detailed discussions of field-plot techniques and data analysis are provided by Gomez and Gomez (1984) and LeClerg et al. (1962). An overview of the general principles will be provided in this chapter.

SOURCES OF VARIATION

The ideal way to compare genotypes would be to grow all of them in exactly the same environment and to measure their characteristics in precisely the same manner. The differences among genotypes in this ideal situation would be due only to variation in their genetic potential; therefore, the best genotype could be chosen without error. This ideal is impossible to achieve under field conditions because of lack of uniformity in the environment to which the genotypes are

261

262   WALTER R. FEHR

exposed. Nevertheless. the use of appropriate field-plot techniques can maximize
the accuracy with which genotypes are compared and selected.

The factors that result in test conditions that are less than ideal can be referred
to collectively as sources of experimental error. They include variation in the
environment to which each genotype is exposed and lack of uniformity in the
measurement of characters. The breeder has opportunities to minimize experi-
mental error by carefully selecting the site to be used for field trials, the cultural
practices used in crop production. the plot size and shape. and the method of
data collection.

Site Selection

Variation in the productivity of the soil is commonly referred to as soil heter-
ogeneity (Fig. 19-1). Causes of soil heterogeneity include variation in soil type,
availability of plant nutrients. and soil moisture. The variation cannot be com-
pletely eliminated. but it often can be minimized by careful selection of the area
in a field where plots will be grown. Soil maps are helpful for understanding
the variation in soil type that is present. Soil types differ in their inherent ability
to retain nutrients and moisture. Entire trials or at least an entire replication
should be grown on a single soil type whenever possible.   .

Visual inspection of a field is important. even when a soil map is available.

Figure 19-1   Example of potential variation in soil productivity in a test area.

HR
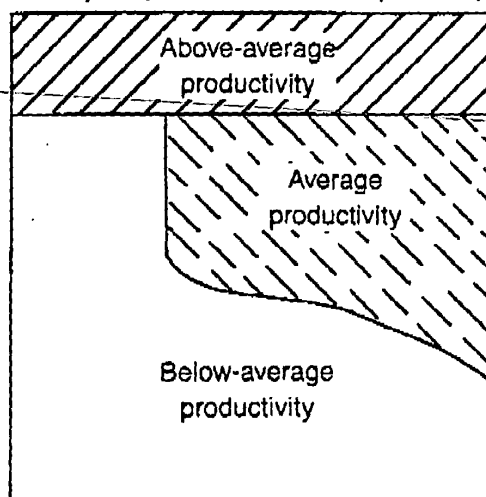
nize

rred
: the
: the
peri-
tural
d of

When a field has been identified a year in advance as a potential test site, it is useful for the breeder to look for variability in productivity of the crop grown in the area. The breeder should note variation in the terrain that may cause water to accumulate more in one place than in another. Differences in soil tillage after harvest of the previous crop may be observed that could result in nonuniformity of the area. Uneven distribution of plant or animal waste on a field should be noted as a potential contributor to variation in the availability of plant nutrients.

Before a site is chosen, information should be obtained on cultural practices that were followed in the production of previous crops, with special attention to the application of chemicals that could influence the crop that the breeder will be evaluating. The residue from herbicides applied for control of weeds in previous crops may cause damage to the crop to be tested. The following quotation from a research article by Thorne and Fehr (1970b) on soybean breeding illustrates the importance of herbicide residue:

leter-
type,
com-
: area
nding
.bility
:ation

> The strains were evaluated at Ames and Kanawha, Iowa, in 1968. . . . At Kanawha, part of the experiment was inadvertently planted in a field treated with atrazine herbicide two years before. All plots in the area were destroyed.

ilable.

Previous cultural practices in a field can be especially important at research stations where crops are rotated from one field to another on a systematic basis. The research conducted on crops previously grown on a field can influence markedly the uniformity of the test site. For example, plots of oats were planted in a field at the Agronomy Research Center of Iowa State University in which soybeans had been planted the previous year. Growth of the oats varied in strips, as if nitrogen fertilizer had been applied unevenly to the field. A review of the previous soybean research revealed that the strips of oats with extra growth coincided with areas where mature soybeans had been cut and left unthreshed. The nitrogen in the soybean seeds in the strips was available to the oats the following year, and caused nonuniformity of nutrient availability in the test site.

: area.

Cultural Practices

Experimental error can be minimized by the use of uniform cultural practices for production of the crop being tested. Chemicals should be applied uniformly to the test site before, during, or after planting. Uneven soil compaction should be minimized during tillage operations. Application of supplemental water by irrigation may reduce variability in soil moisture. Weed control should be uniform; most breeders try to eliminate all weeds during the growing season to avoid experimental error caused by differential weed competition.

The development of equipment specifically designed for planting, managing, and harvesting research plots has permitted breeders to grow plots more efficiently. The emphasis in the design and use of any equipment must be on the uniformity with which genotypes are handled.

Plot Type

Experimental error increases whenever interplot competition causes the perfo
mance of a genotype in one plot to be altered by the performance of genotyp
in adjacent plots. Interplot competition results primarily from intergenotyp
competition, which is the differential ability of genotypes to compete with eac
other. Interplot competition is more important for the evaluation of some cha
acters than for others. It is only through appropriate experimentation that a pl
type can be identified that will provide reliable information for the character c
interest.

The effects of interplot competition can be avoided by the use of plots wit
multiple rows in which only plants in the center rows are evaluated (Fig. 19-2.
In plots with three or more rows. the outermost rows are designated as the borde
or guard rows. The function of the border rows is to prevent plants in adjacer

Figure 19-2   Illustration of bordered row plots with different cultivars desig
nated as ●. ○ and ☐. (Courtesy of Fehr, 1978.)

Bordered row plots - equal row spacing

Three-row	Four-row	Five-row
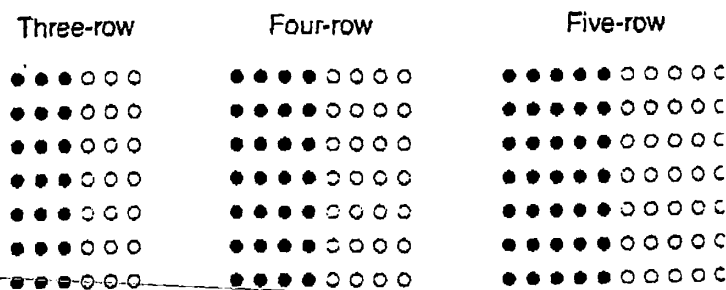
Bordered row plot - unequal row spacing
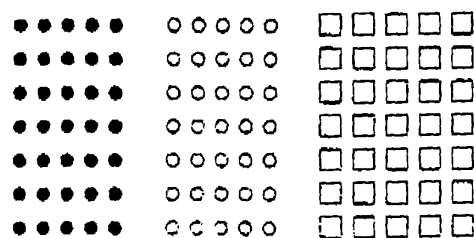
FIELD-PLOT TECHNIQUES   265

plots from influencing the performance of plants in the center of the plot. Each bordered plot can be considered a miniature field that is unaffected by neighboring fields. The spacing between plots can be greater than the within-plot spacing to facilitate the movement of equipment, particularly when narrow rows are utilized.

It would be ideal if bordered plots could be used for the evaluation of all characters that are influenced by interplot competition. That ideal is difficult to achieve when thousands of genotypes are being evaluated. Bordered plots require seed and land that do not directly provide data for a genotype. Borders take up two-thirds of the seed and land area for three-row plots and one-half for four-row plots. The cost and availability of seed and land often necessitate restriction of the use of bordered plots to the evaluation of genotypes that are being given final consideration for release as cultivars.
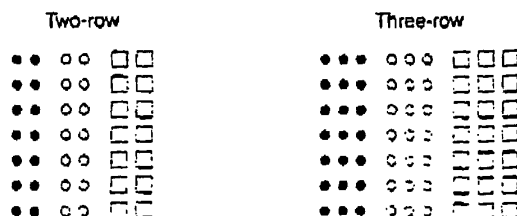
Interplot competition can be reduced, but not eliminated, with unbordered plots of two or more rows, all of which are used to evaluate a character (Fig. 19-3). A genotype in a single-row plot is subjected to interplot competition on both sides. Interplot competition is reduced by one-half in plots with two rows, two-thirds with three rows, three-fourths with four rows, and four-fifths with five rows. The estimated reduction of interplot competition with increasing numbers of rows is based on the fact that each row of a plot must compete on two sides. The border rows are each subjected to interplot competition on one side

Figure 19-3   Illustration of unbordered row plots with different cultivars designated as ●, ○, and ▢. (Courtesy of Fehr, 1978.)

Unbordered row plots - equal row spacing

Single-row   Two-row   Three-row
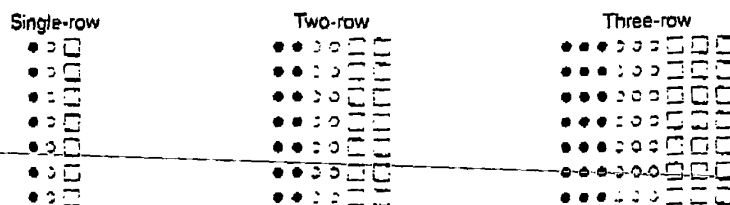
Unbordered row plots - unequal row spacing

Two-row   Three-row

266   WALTER R. FEHR

but not on the other. Any rows within the two border rows are protected from interplot competition. This can be expressed as

$$\text{Reduction in interplot competition compared with single-row plot} = \frac{(\text{number of rows per plot} \times 2 \text{ sides}) - 2 \text{ sides}}{\text{number of rows per plot} \times 2 \text{ sides}}$$

$$\text{Two-row plot} = \frac{(2 \times 2) - 2}{2 \times 2} = 1/2$$

$$\text{Three-row plot} = \frac{(3 \times 2) - 2}{3 \times 2} = 2/3$$

The amount of interplot competition also can be reduced by increasing the spacing between rows of adjacent plots. Interplot competition in soybeans was evaluated with five cultivars grown in single rows spaced 100, 75, 50, and 25 cm apart (Gedge et al., 1977). The average effect of interplot competition on seed yield was 2.6 percent for the 100-cm spacing, 5.3 percent for 75 cm, 8.0 percent for 50 cm, and 17.6 percent for 25 cm.

A combination of increased row spacing between plots and a large number of rows can minimize interplot competition in unbordered plots. In the soybean example of the preceding paragraph, the average change in yield for single-row plots spaced 100 cm apart was 2.6 percent. The percentage theoretically would be reduced to 1.3 percent for two-row plots and to 0.9 percent for three-row plots. Rows within a plot are not subjected to interplot competition; therefore, the spacing between rows within a plot can be less than the spacing between adjacent plots. Figure 19-3 illustrates a two-row plot in which the spacing between plots is wide enough to minimize interplot competition and the spacing within the plot is reduced to minimize the land area required for each plot.

Some breeders plant one cultivar as a common border between one- or two-row plots. In barley, a lodging-resistant cultivar is used as a common border to prevent genotypes with lodging susceptibility from falling on genotypes in adjacent plots, thereby causing them to lodge unnaturally. The use of a common border has been evaluated as a means of eliminating intergenotypic competition between plots for seed yield and other quantitative characters. The results of the research indicate that a common border can reduce but not eliminate interplot competition (Thorne and Fehr, 1970a). The average interplot competition for seed yield among four soybean cultivars in single-row plots spaced 50 cm apart was compared with competition of the cultivars when a common border was used (Gedge et al., 1977). Interplot competition averaged 11.0 percent in single-row plots and 8.3 percent in plots with a common border.

Plot Size and Shape

The size of plots used to evaluate genotypes varies with the character being evaluated, the amount of experimental error that is considered acceptable for

measuring a character, the experimental design, and the growth characteristics of the crop. Plots vary in size from those for a single plant that is harvested by hand to those that are wide and long enough to be harvested with the same equipment used by farmers for commercial production.

Single-Plant Plots. Individual plants commonly are evaluated in segregating populations. There is no replication of the individuals, unless vegetative propagation of clones is possible. The spacing among plots varies with the crop species involved. Gardner (1961) spaced individuals 50 by 100 cm apart when selecting for yield in maize. Burton (1974) spaced plants of a population of Pensacola bahiagrass 60 by 60 cm apart when conducting recurrent phenotypic selection for forage yield. Burton and Brim (1981) used a 46 by 46 cm spacing among soybean plants for selection of oil composition in the seed.

Single-plant plots are used for the replicated evaluation of experimental lines or cultivars by the honeycomb field design (Fasoulas, 1979). The number of plants evaluated for a line is equal to the number of replications in the experiment. Fasoulas (1981) indicated that 100 single-plant plots (replications) per line would provide satisfactory results. The plots of the lines in a test are organized in a systematic manner to permit comparison of a plant of one line with adjacent plants of other lines (Fig. 19-4). The honeycomb design has not been adopted by plant breeders for replicated evaluation of lines because it requires more labor and is less amenable to mechanization than microplots or conventional row plots.

Multiple-Plant Plots. The evaluation of experimental lines or cultivars by plant breeders is usually done in plots containing two or more plants. Plot size varies from small microplots consisting of a hill or short row to a plot with one or more rows several meters in length.

Microplots. Microplots are used to minimize the amount of seed or land required to evaluate a group of lines. In an unbordered microplot, the effects of interplot competition must be considered when determining an appropriate distance among plots (Fig. 19-5). For oats, hill-plots spaced about 30 by 30 cm apart have been used (Frey, 1965), while for soybeans, a spacing of about 1 by 1 m is more common (Gurland and Fehr, 1981).

The number of plants in a microplot differs among crops. A planting rate of 30 seeds per hill is satisfactory in oats (Frey, 1965), while a rate of 12 seeds per hill is used for soybeans (Garland and Fehr, 1981). When short rows are used as microplots, the plant density is comparable to that of larger row plots.

There is a lack of agreement among plant breeders concerning the effectiveness of microplots for evaluation of agronomic characters, particularly seed yield. Breeders who use microplots indicate that they are useful for eliminating inferior lines during the first year of yield evaluation. Lines with acceptable performance in microplots are evaluated in conventional row plots during subsequent years of testing, to identify those that merit release as cultivars (Frey,

Grid design

```
X   X │ X   X │ X   X │ X   X
X   X │ X   X │ X   X │ X   X
X   X │ X   X │ X   X │ X   X
X   X │ X   X │ X   X │ X   X
X   X │ X   X │ X   X │ X   X
X   X │ X   X │ X   X │ X   X
X   X │ X   X │ X   X │ X   X
```
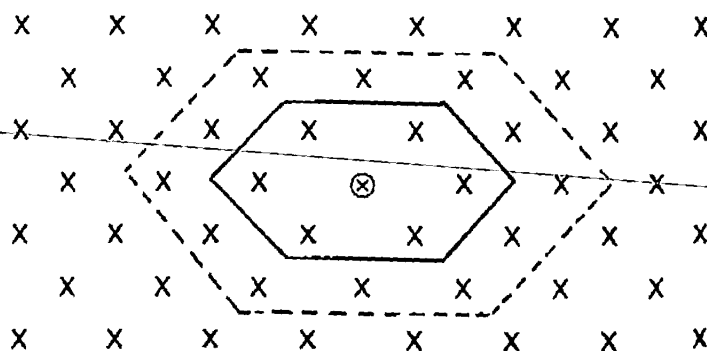
Honeycomb

Figure 19-4   Grid and honeycomb design to select individual plants in a population. For the grid design. plants are divided into blocks and the best ones chosen from each (Gardner. 1961). For the honeycomb design. the plant at the center of the hexagon. ⊗. is compared with every other plant within the hexagon (Fasoulas. 1979). A plant is chosen only if it is superior to every other plant in the hexagon. The hexagons outlined represent two different selection intensities.

268

HILL PLOTS

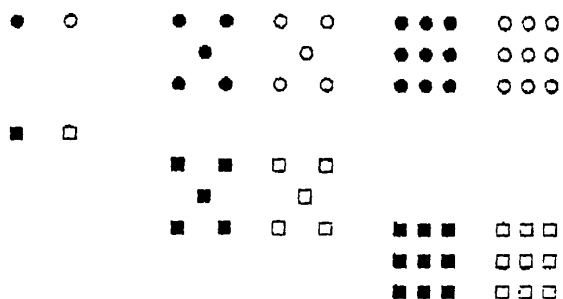Unbordered   Bordered

Single-hill   Five-hill   Nine-hill

Figure 19-5 Illustration of hill plots with different cultivars designated as ☐. ◌. ●. and ■ (Fehr. 1978).

1965: Garland and Fehr. 1981). The advantages of microplots compared with conventional row plots for the first year of yield testing are that less land is required per plot and that enough seed for replicated tests can be obtained from a single plant. which eliminates a season for seed increase. Breeders who do not use microplots are concerned about the reliability of yield data obtained from them. The coefficients of variability for microplots generally are about one and one-half to two times larger than for conventional row plots.

Row Plots. Row plots are used by virtually all plant breeders for replicated testing of genotypes. The overall plot size is determined by the number of rows. the spacing between rows. and the row length.

Single-row plots of 1 to 2 m in length are widely used for the visual evaluation of characters. Many breeders evaluate lines on the basis of their appearance in small unreplicated plots. and advance the desirable ones to replicated tests the following season. Visual selection and seed increase commonly are accomplished with the same plot.

A plot used to evaluate the yield of lines for the first time often is smaller than that employed for advanced stages of evaluation. For advanced yield tests. the breeder attempts to use a plot size that approaches or equals the dimensions considered optimal for the crop species involved. Optimum plot size is the minimum land area required to measure a character with an acceptable level of experimental error.

Optimum plot size can be determined by the use of data from a uniformity trial (Cochran. 1937). A single cultivar is planted as a solid stand. without alleys.

in an area representative of that used for yield evaluation. The cultural practices used to produce the crop are the same as those used for yield trials. The area is subdivided into small units, and the seeds or plants from each unit are harvested and weighed separately. Experimental error associated with plots of different size can be determined by making various combinations of the small units.

Optimum plot size also is determined through practical experience. The breeder often will experiment with plots of different size to find the smallest one that has an acceptable level of experimental error. Breeders often do not agree on what they consider acceptable experimental error: consequently, an optimum size for one person may not be optimum for another.

Plot width generally is determined by considerations other than the relationship of shape to experimental error. The primary factors are the number of rows required to minimize or avoid interplot competition and the width of the planting and harvesting equipment that is available. Plot width influences the percentage of land area that must be devoted to alleys between plots. Long, narrow plots require a lower percentage of alley space than do wide, short plots. This advantage is offset in bordered plots because the percentage of land area devoted to border rows decreases as the number of rows per plot increases.

Plot length provides flexibility for plot size. Before calculators and computers became readily available, row length in the United States was varied to obtain a plot size that was a fraction of an acre (one-tenth, one-twentieth, etc.) to simplify the conversion of plot yields to yields per acre. With use of computers for data summarization and analysis, this is no longer necessary.

Data Collection

The experimental error associated with the evaluation of a character is influenced by measurement errors during data collection. For characters evaluated visually, experimental error occurs whenever the data collector fails to give an identical rating to plots with an identical appearance. Reliability of the evaluation can be established readily by rating a series of plots at different times and comparing the ratings. It is essentially impossible to give visual ratings without error; therefore, the breeder must decide when the error is acceptable and when it is so large that genetic differences will be masked.

Some characters can only be evaluated efficiently with the use of an appropriate machine or instrument. Experimental error can occur because of failure to prepare a plot properly for measurement, of not obtaining a representative sample of the plot for evaluation, of using nonuniform procedures for sample preparation, and of failure of the machine or instrument to operate properly.

Preparation of a plot for data collection may begin before planting. For experimental error to be reduced, the seeds or plants of every genotype used for planting must be treated equally. If seeds or plants of genotypes to be compared

do not come from a common environment, environmental error may result. Lint yield and seedling vigor of a cotton cultivar were found to differ in plots grown from seeds obtained from different locations (Peacock and Hawkins, 1970). Seed source also has been shown to influence seed yield of soybeans (Fehr and Probst, 1971.)

In some crop species, uniformity of plant density among plots can be important in minimizing experimental error. With maize, it is a common practice to thin yield test plots to a uniform stand soon after seedling emergence. Thinning is not considered necessary with some crop species, particularly those that have the ability to branch or tiller in response to low plant density, such as barley and wheat. It also is a common practice with crops such as maize to record the number of plants per plot immediately before harvest. The yield of the plots is adjusted for plant density by an analysis of covariance, to minimize experimental error in the comparison of genotypes.

When a blank alley is used at the end of row plots, the end plants generally are more productive than those growing in the center of the plot. When end plants are harvested, yield of the plot is inflated in comparison to the yield obtained from plants growing in the center of the plot. This inflation will prevent a direct comparison of plot yields with those expected in a normal commercial planting, unless an appropriate adjustment is made for all plots. The adjustment may be made by considering the alley as part of the plot area; therefore, plot length is the distance from the center of one alley to the center of the next, instead of the distance between plants at opposite ends of a row. For example, if the length of row containing plants is 5 m and the alley is 1 m wide, the plot length for computing plot area is considered to be 6 m.

The yield inflation by end plants in a plot does not contribute to experimental error unless genotypes in a test do not respond similarly to the space in the alley. The experimental error associated with differential response of genotypes to an alley can be minimized by adjusting yields according to characteristics of the genotypes that influence this response. The end plants of soybean genotypes with late maturity give a greater yield inflation than do genotypes of early maturity. Values have been developed with which to adjust plot yields for maturity of soybean genotypes (Wilcox, 1970). More commonly, comparisons among soybean genotypes are restricted to those of similar maturity, unless plots are end-trimmed before harvest.

The only way to eliminate yield inflation by end plants is to remove the plants before harvest. This procedure, referred to as end-trimming, is a standard procedure with some crops. The end plants are removed late enough in plant development that the remaining plants in the plot cannot take advantage of the extra space. The length of row removed from each end of the plot must be long enough to include all plants that have benefited from the space provided by the alley. In soybean, 0.6 m is removed from each end of the plot (Wilcox, 1970).

The problem of a blank alley is minimized in some crops by planting the

272   WALTER R. FEHR

alley with rows of a single genotype perpendicular to the test plots. The result
is that the plants at the end of a plot must compete with plants in the alley, and
thus their yield may not be inflated as much as is the case with a blank alley.
Plants in the alley are removed immediately before the plots are harvested.

EXPERIMENTAL DESIGNS

The arrangement of genotypes in a field experiment is referred to as the exper-
imental design. Some of the designs utilized to compare genotypes are common
to research in many disciplines. Others have been developed to deal with the
problem of comparing a large number of genotypes as inexpensively as possible.
The experimental designs used for the initial evaluation of a large number of
genotypes often differ from those used in the advanced stages of testing a few
select genotypes. Alternative designs will be considered here for comparison of
single plants, unreplicated genotypes in multiple-plant plots, and replicated
genotypes.

Single-Plant Selection

The first evaluation step in the development of a cultivar generally is the selection
of individual plants from a population. Individual plant selection also is employed
in population improvement by recurrent phenotypic selection.

   When single-plant selection in a population is for characters with a high
heritability, the plants generally are grown in a random order and those with
desirable characteristics are selected. Cultivars may be grown in adjacent plots
to serve as standards with which to evaluate single plants. Date of flowering,
plant height, time of maturity, and certain types of pest resistance are examples
of characters for which single plants are selected without any predetermined
arrangement of the individuals. They represent characteristics that are not strongly
influenced by environmental variation.

   Single-plant selection in a population grown in a relatively large land area
can be hampered seriously by soil heterogeneity for characters with a low her-
itability, such as seed or plant yield. Figure 19-1 illustrates variation in soil
productivity in an area where a population of plants may be grown. If plants
with the highest yield are selected regardless of their location in the field, those
in the area of above-average productivity will be favored. A plant with outstand-
ing genetic potential that is located in the area with below-average productivity
may be discarded. Two experimental designs are available that minimize the
effect of soil heterogeneity by comparing plants that are most adjacent to each
other.

Grid Design. Gardner (1961) proposed that the land area on which a population
of individual plants is grown can be subdivided into blocks or grids of a limited

area (Fig. 19-4). Plants within each block are compared with each other, and the superior ones are selected. Comparisons are not made between plants from different blocks. This experimental design has been well accepted by plant breeders, particularly those conducting recurrent phenotypic selection for yield or other characters with a low heritability.

Honeycomb Design. Fasoulas (1973) developed a honeycomb design for selecting individual plants in a population (Fig. 19-4). Five aspects of the design and its implementation are unique. (a) Seeds or clones are spaced equidistantly from each other in a hexagon pattern. The name of the design was chosen because the hexagon patterns resemble a honeycomb of bees. (b) Plants are spaced far enough apart that they do not compete with adjacent individuals. At the appropriate spacing for a species, a missing plant does not influence the performance of adjacent individuals, because each plant already has sufficient space in which to develop to its full potential. (c) Homogeneous check cultivars can be included for comparison, if desired. Every plant of the check is compared with a different group of plants in the population. (d) The size of the hexagon used to select single plants determines the selection intensity in the population. The effect of soil heterogeneity is minimized because only those plants within the area of the hexagon are compared. (e) Every plant in the population is evaluated by placing it in the center of the hexagon. A plant is chosen only if it is superior to every other plant in the hexagon. By moving the hexagon, every plant is compared with a different group of plants in the population.

Comparison of the Grid and Honeycomb Designs. Both the grid and honeycomb designs reduce the problem of soil heterogeneity in the selection of characters of low heritability. In a comparison of the designs, the advantages of one are the disadvantages of the other, and vice versa.

There are three primary advantages of the grid design.

1. The spacing of plants does not have to be in a precise pattern. This facilitates the use of conventional plot equipment for planting and cultivation. Mechanized planting of the honeycomb design would require specialized equipment.
2. Selection intensity can be varied by altering the number of plants in a block and the number of plants selected. Only certain selection intensities are possible with the honeycomb design.
3. Use of a defined area for each block facilitates visual comparison of plants for selection. It is possible to compare plants within a block visually and collect data only from those with the best potential. Use of the moving hexagon for the honeycomb design makes it impractical to compare each plant with appropriate ones in its hexagon; therefore, data must be recorded for every plant, except those that are obviously inferior.

The honeycomb design has two advantages compared with the grid design.

274   WALTER R. FEHR

1. Homogeneous check cultivars can be included to permit comparisons of individual plants with a standard. When one-seventh of the plants are a check, they can be arranged so that every plant in the population can be compared with a check plant. To provide adjacent plants of one check cultivar in a grid system, one-third of the area would have to be devoted to the check.

2. More than two check cultivars can be included readily in hexagons of 19 or more plants. Use of two or more check cultivars in the grid system would require that a large fraction of each block be devoted to check plants.

Unreplicated Evaluation with Multiple-Plant Plots

Plant breeders routinely conduct visual selection among lines in unreplicated plots for maturity, disease resistance, standability, and other characters of high heritability. Evaluation for yield in a single replication has been used to a limited extent to eliminate inferior lines before initiation of expensive replicated tests. With a single replication, each line is compared once with check cultivars or other lines to determine its genetic potential. A number of different arrangements are available for estimating the genetic potential of lines. One method is to compare each line with a common check cultivar (Baker and McKenzie, 1967). Figure 19-6 represents a hypothetical example of the yield of six lines in a single replication. In the figure, the yield of each line is expressed as a percentage of the yield of the check cultivar immediately adjacent to it.

Another alternative is to express the yield of each line as a percentage of the weighted average of the adjacent check plot and of the check plot two plots removed. The purpose for using a weighted average is to minimize the potential problem caused by an unusually poor yield of a check plot. In Fig. 19-6, the check cultivar adjacent to lines B and C has a much lower yield than other check cultivars. This results in an extremely high percentage for lines A and B. The weighted average of check cultivars could be computed as

($\frac{2}{3}$ × yield of adjacent check) + ($\frac{1}{3}$ × yield of check two plots removed)
                                                = weighted average of check cultivars

The percentage yield of each line is computed as

$$\text{Line A} = \frac{59}{(\frac{2}{3} \times 55) + (\frac{1}{3} \times 39)} \times 100 = 119$$

$$\text{Line B} = \frac{70}{(\frac{2}{3} \times 39) + (\frac{1}{3} \times 55)} \times 100 = 158$$

$$\text{Line C} = \frac{53}{(\frac{2}{3} \times 39) + (\frac{1}{3} \times 48)} \times 100 = 126$$

$$\text{Line D} = \frac{51}{(\frac{2}{3} \times 48) + (\frac{1}{3} \times 39)} \times 100 = 113$$
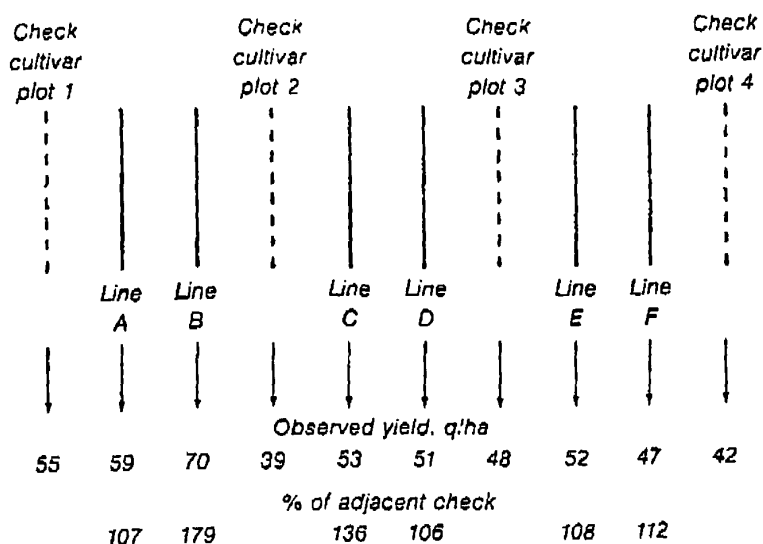
FIELD-PLOT TECHNIQUES   275

Figure 19-6   One possible arrangement of lines in a single-replication test. The performance of each line is computed as a percentage of the performance of the common check cultivar adjacent to it. Line B would be considered the superior one.

$$\text{Line E} = \frac{52}{(\frac{2}{3} \times 48) + (\frac{1}{3} \times 42)} \times 100 = 113$$

$$\text{Line F} = \frac{47}{(\frac{2}{3} \times 42) + (\frac{1}{3} \times 48)} \times 100 = 107$$

Another method used to compare genotypes in single replications is the moving mean (Mak et al., 1978; Townley-Smith and Hurd, 1973). Each genotype is compared with adjacent test genotypes, not with a check cultivar.

The disadvantage of single-replication tests is that the breeder has only one plot value with which to assess the genetic potential of a line. If by chance a line is placed on a plot of soil with above-average productivity, relative to that of plots with which the line is compared, it will seem to be genetically superior, even though it may not be. In replicated tests, the breeder will have more than one plot with which to evaluate each line. For this reason, single replications are not commonly used for yield evaluation.

Replicated Tests

Two or more independent comparisons of lines in a test provide a means of estimating whether variation in performance among lines is due to differences in genetic potential or to environmental variation. Each comparison is as rep-

276   WALTER R. FEHR

llcation. Replication can be accomplished by growing two or more plots of each
line at one or more locations or one plot at each of two or more locations or
years.

Randomization. One important consideration in the arrangement of genotypes
within each replication is the degree of randomization. From a statistical view-
point, randomization of entries is required to obtain a valid estimate of experi-
mental error. To fulfill the requirement, each entry must have an equal chance
of being assigned to any plot in a replication and an independent randomization
is required for each replication.

Plant breeders understand the importance of randomization and consider it
the ideal procedure for comparison of genotypes. They know that any experiment
designed to estimate components of variance must be randomized. There are
circumstances. however. in which plant breeders do not use complete random-
ization for the comparison of genotypes. Genotypes with similar characteristics
may be planted next to each other to reduce interplot competition in unbordered
plots. A nonrandom arrangement of genotypes among replications may be used
to facilitate selection of genotypes before harvest.

Nonrandom Arrangements of Genotypes. Any discussion of nonrandom arrange-
ments of genotypes can be misinterpreted because it may imply that randomi-
zation is not an important principle. To avoid such misinterpretation. it should
be stated again that nonrandomization should only be considered when resources
are not adequate to make randomization feasible. The discussion of nonrandom
arrangements will include the reasons for their use. their disadvantages. and the
ways procedures can be modified to permit effective randomization.

Nonrandomization Among Replications. It is common to delay replicated tests
for yield until genotypes have been visually selected in unreplicated plots for
characteristics-such-as-lodging, height, and maturity. To reduce the length of
time for cultivar development, the season for evaluation in unreplicated plots
can be eliminated by growing genotypes in replicated plots, visually selecting
those with desirable characteristics. and harvesting only the plots of selected
genotypes for yield evaluation (Garland and Fehr. 1981). When visual selection
is based on the performance of genotypes in all of the replications. it is necessary
to evaluate each plot. summarize the data. make the selections. and identify the
plots of selected genotypes that should be harvested. The length of time between
plot evaluation and harvest may be only a few days when characteristics of
interest are not expressed until plant maturity. If several thousand genotypes are
randomized in two or more replications. summarization of data and identification
of plots to be harvested can be difficult or impossible to accomplish in only a
few days. The use of the same arrangement of genotypes in each replication
makes the job practical.

When genotypes are in the same position within each replication, the data
for plots of each genotype are recorded in adjacent columns (Fig. 19-7). Sum-

Nonrandom

Plot	Entry	Replication		
		1	2	3
1	1			
2	2			
3	3			
4	4			
5	5			
6	6			

- -

Random

Plot	Entry	Replication 1
1	4	
2	1	
3	6	
4	3	
5	5	
6	2	

Plot	Entry	Replication 2
1	5	
2	4	
3	2	
4	1	
5	6	
6	3	

Plot	Entry	Replication 3
1	2	
2	6	
3	3	
4	5	
5	1	
6	4	

Figure 19-7   Field book pages for recording the data of genotypes grown in three replications. Nonrandom arrangement of genotypes involves one page, whereas a random arrangement involves three separate sections on one or more pages.

277

278   WALTER R. FEHR

marization of data is complete as soon as the last plot is rated. Genotypes with
undesirable characteristics in one or more replications can be identified and
discarded. The plots of desirable genotypes are readily identified for harvest
because they are in the same position in each replication.

The disadvantages of nonrandomization relate to the fact that the same ge-
notypes are always adjacent to each other, which can have negative effects on
the comparison of genotypes.

1. In unbordered plots, intergenotypic competition can bias the performance
   of genotypes more seriously in a nonrandom than in a random arrange-
   ment. When a poor competitor is bordered by a good competitor, yield
   of the poor competitor can be reduced and that of the good competitor
   increased in every replication. There is no opportunity for a genotype to
   occur next to others with a more similar competitive ability.
2. In unbordered plots, a genotype that dies or is unusually weak in all
   replications can prevent the accurate evaluation of adjacent genotypes.
   The performance of adjacent genotypes would never be tested in repli-
   cations where they were next to healthy genotypes.
3. No unbiased estimate of experimental error can be obtained.

The need to use nonrandomization of genotypes among replications can be
avoided by improving the efficiency of procedures for data summarization and
evaluation. An efficient procedure would include the use of a computer. Data
would have to be entered rapidly into the computer, possibly by entering plot
data into an electronic recorder in the field and electronically transferring the
information to the computer. Computer programs would be needed to summarize
the data and make selections on the basis of standards established by the breeder.
Plot identification information for selected genotypes would have to be provided
for harvest.

Grouping Similar Genotypes Within Replications. The evaluation of genotypes
in unbordered plots can be hampered by bias from intergenotypic competition.
Plant characteristics that often contribute to intergenotypic competition in a crop
include such factors as differences in height and time of maturity. To reduce
intergenotypic competition, genotypes with similar characteristics may be grouped
within replications. The position of each genotype may be varied from one
replication to the next. This procedure, sometimes referred to as restricted ran-
domization, has the advantage of reducing the effects of intergenotypic com-
petition in unbordered plots. The primary disadvantage is that all genotypes in
a test cannot be compared with the same level of confidence. Genotypes within
a group are spaced closer to each other than genotypes in different groups and
are less affected by environmental variation among plots.

The use of bordered plots eliminates the need for grouping genotypes. The
performance of genotypes in plots is not influenced by intergenotypic compe-

FIELD-PLOT TECHNIQUES   279

tition; therefore, randomization is practical. An increase in land, seed, and other resources will be needed for replacement of unbordered plots with bordered ones.

Experimental Designs for Replicated Tests. The arrangement of genotypes in replicated tests involves primarily the use of either the randomized complete-block design or incomplete-block designs. The Latin square is used only in special circumstances when the number of entries is small (Cochran and Cox, 1957). The honeycomb design can be used for replicated testing but is considered too difficult to implement for a large number of lines (Fasoulas, 1981).

The differences between the randomized complete-block and incomplete-block designs relate to their ability to account for environmental variation within a replication. The two types of design differ in restrictions on the size of a replication, randomization procedures, analysis of data, and comparisons among genotypes.

The terms complete-block and incomplete-block refer to the arrangement of genotypes in an experiment (Fig. 19-8). A block and a replication are equivalent in a randomized complete-block design. A block contains all of the genotypes in the test and is considered complete. Genotypes are divided into more than one block within each replication of an incomplete-block design. The blocks are considered incomplete because they contain only part of the genotypes. A number of different types of incomplete-block designs are available (Cochran and Cox, 1957). The most common types used in plant breeding are referred to as lattices. In a lattice design, a replication is divided into blocks that collectively contain all the genotypes in a test (Fig. 19-8).

The incomplete-block designs are intended to provide more control over environmental variation within a replication than is possible with the complete-block design. The ideal situation for genotype evaluation would be to test each genotype in the same plot, thus avoiding any environmental variation caused by differences in soil fertility, moisture, and other factors within a field. This is not possible, so the next best approach is to adjust the performance of each genotype according to the relative productivity of the plot in which it is evaluated. If one plot has better fertility and moisture than the average-for-all-plots in a replication, the performance of a genotype in that plot will be adjusted downward. A genotype in a plot with lower productivity than the average will have its performance adjusted upward.

Although individual plot adjustments are not possible, the lattice designs permit the performance of a genotype to be adjusted upward or downward according to the productivity of the blocks in which it was grown. The randomized complete-block design does not divide the replication into smaller units and is not able to adjust the performance of a genotype for environmental variation within replications.

The effectiveness of the lattice design in accounting for environmental variation within replications depends on the pattern of variation. Figure 19-9 shows two replications with variation in soil productivity. The soil productivity in

280   WALTER R. FEHR

Block	Replication 1					
1	1	2	3	4	5	6
2	7	8	9	10	11	12
3	13	14	15	16	17	18
4	19	20	21	22	23	24
5	25	26	27	28	29	30
6	31	32	33	34	35	36
7	37	38	39	40	41	42

	Replication 2					
1	7	13	19	25	31	37
2	1	14	20	26	32	38
3	2	8	21	27	33	39
4	3	9	15	28	34	40
5	4	10	16	22	35	41
6	5	11	17	23	29	42
7	6	12	18	24	30	36

|	Replication 3					
1	12	17	22	28	33	38
2	2	13	24	29	35	40
3	4	9	20	25	36	42
4	6	11	16	27	32	37
5	1	7	18	23	34	39
6	3	8	14	19	30	41
7	5	10	15	21	26	31

Figure 19-8 Lattice design for an experiment with 42 entries and three rep-
lications. (Adapted from Cochran and Cox, 1957.) For a randomized complete
block design, there are no blocks within a replication and the entries are assigned
at random to the 42 plots.

replication 1 increases from left to right. The blocks of the lattice design ar
arranged in a pattern that effectively measures the variation, as evidenced b;
differences in the mean for each block. The variation in soil productivity i
replication 2 does not fit a consistent pattern. Much of the variation occurs withi
blocks, and the mean performance of the blocks is relatively similar. The latti

FIELD-PLOT TECHNIQUES   281

Replication 1   Replication 2

40 39 45 54 58 67 70 72   52 60 54 57 51 59 55 56
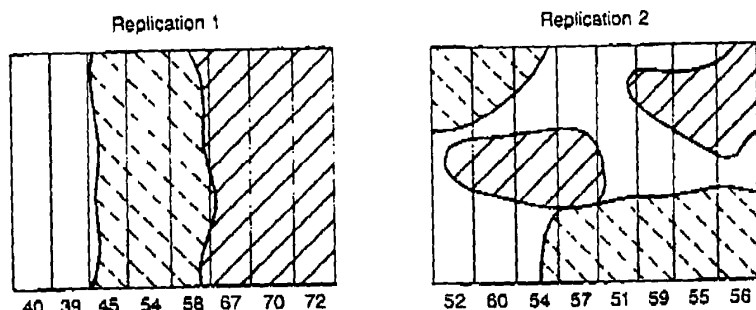
Mean performance of entries in a block

= soil with high productivity

= soil with average productivity

= soil with low productivity

Figure 19-9 The effect of the pattern of variation in soil productivity on the effectiveness of the lattice design in accounting for environmental variation within a replication. The lattice would be more effective in replication 1 than in replication 2.

design-cannot adjust for differences in productivity within a block; therefore, it would not be as effective in replication 2 as in replication 1.

The effectiveness of the lattice design compared with the randomized complete-block is expressed as relative efficiency. Relative efficiency is computed as a ratio of mean squares for experimental error of the two types of design.

$$\text{Relative efficiency} = \frac{\text{mean square for error of lattice}}{\text{mean square for error of randomized complete-block}} \times 100$$

The ratio is used to determine the number of replications that would have to be used with the randomized complete block to achieve a precision in detecting differences among the means of genotypes equal to that with a lattice design. A relative efficiency of 150 percent indicates that 50 percent more replication would have been needed with a randomized complete-block design than with a lattice.

The two types of design differ in the flexibility that is possible in a test. The randomized complete-block can accommodate any number of genotypes or replications. The lattice design requires that a specified number of genotypes and replications be included. For example, no lattice design can be used with 44, 58, or 74 genotypes. There is no restriction in a randomized complete-block for the length and width of a replication. For example, a test with 72 entries could be planted 8 plots long by 9 plots wide or 6 plots long by 12 plots wide. The shape of replication for a particular number of genotypes in a lattice is not as flexible. A test with 72 entries could be planted 8 plots long by 9 plots wide, not 6 plots long by 12 plots wide.

The randomization of an experiment and statistical analysis of data are more complex for a lattice than for a randomized complete-block. This can be important if the work is done by hand, but not if done by computer. Computer programs are available that will readily accommodate either type of design.

EQUIPMENT FOR EFFICIENT EVALUATION OF GENOTYPES

The efficient evaluation of a large number of genotypes is important for genetic improvement. Plant breeders have been actively involved in the development of equipment that permits them to evaluate more genotypes with equal or greater quality than was previously possible. The equipment ranges from simple hand devices to sophisticated computers.

Each crop has unique characteristics that influence the type of equipment used. Even for a certain crop, breeders differ as to the type of equipment they consider most desirable. Here only a small sample of available equipment will be used to illustrate how large numbers of genotypes are evaluated by plant breeders.

Preparation of Seed for Planting

The main steps involved in preparing a field experiment include packaging the seed and placing it in the proper arrangement for planting. Computers can be used to randomize entries and assign plot numbers. The computer system can print an adhesive label for each packet of seed to be packaged. The label contains the plot number, the entry number, and other information of value to the breeder. The plot and entry information also can be printed on pages used to record data in the field. The same work can be done by hand, but would require a large amount of labor and would be more subject to human error.

Seed is counted by hand or by electronic counting devices. If the number of seeds for a plot is large and precise numbers are not required, the seeds may be measured by volume.

Planting

Rapid planting of plots can be accomplished with engine-driven planters. Multiple-row plots may be planted from a single packet when each row does not require the exact same number of seeds. The seed is passed through a divider that separates the seed into a fraction for each row. The divider may be a powered spinning device or a gravity system.

The planter can move through the field without stopping. Seed for a row is placed in a container above a planting cone. When the row is to be planted, the container is lifted and the seed drops onto the planting cone. Two types of cones are used to distribute seed along the row. For one type, the base turns and carries the seed to the outlet. There it is knocked from the base by a stationary plate, falling through the outlet to the soil. This type of cone is used for relatively small seeds that do not roll easily, such as barley. The second type has fins mounted on the center cone. The seed falls onto a stationary base and is dragged by the fins to the outlet. The fins are well suited to relatively large seeds, particularly those that have a tendency to roll easily, such as maize and soybean. The length of a plot is a function of the distance traveled by the planter before all the seed has left the cone. At a constant ground speed, a cone must turn faster for short rows than for long rows. Adjustment of the speed of the cone rotation can be accomplished readily by several mechanical systems.

While the seed for one plot is being planted, the seed for the next plot is put in the container above the cone. There are a number of ways to determine when the container should be lifted to begin a plot. One way is to mark the beginning and end of each plot in the field before planting starts. When the planter reaches the beginning of a plot, the operator lifts the containers manually or electronically. The advantage of this procedure is that the location of each plot can be identified as soon as planting is complete. The second way is to use a cable extended across the field that has knobs spaced along it. The spacing between knobs is equal to the length of the plot and the alley. For plots that have rows 5 m long with a 1 m alley between them, the knobs would be spaced 6 m apart. As the planter passes by the cable, the knobs signal when the container should be lifted manually, or it activates an electronic tripping device. The cable is moved after each pass across the field. Use of the cable saves time at planting by eliminating the need to mark the start and end of plots manually.

Weed Control

Weed control is accomplished by the use of chemicals, cultivation, and hand weeding. The chemicals generally are those applied for weed control in commercial production of the crop. Cultivation equipment may be especially designed for use in research fields or may be the same equipment used commercially.

Preparation of Plots for Harvest

Trimming of plots to a constant length before harvest is done manually or with specialized equipment. Plots of small grains generally are trimmed to a constant length early in the season when the plants are about 30 cm tall. A rototiller or mower is passed along the end of each plot to kill the unwanted plants. The rototiller may be mounted on a tractor or may be a self-propelled unit that a person walks behind. Plots of soybean can be cut to a constant length with rotary mowers before seed filling begins. Two mowers are attached to a pipe so that they are separated by a distance equal to the desired plot length. and are driven perpendicular to the length of the rows.

Harvest

The most common type of harvester for the measurement of forage yield in the United States is a self-propelled flail chopper. The machine cuts the plants with a rotating flail that throws the cut portion into a collection point behind the driver. The plant material for a plot may be collected in a plastic container and weighed on a stationary scale set up in the field. To eliminate the labor required to use containers. an electronic scale can be mounted on the machine. The plant material is weighed and then it is discarded into a wagon.

The harvest of plots for their seeds is conducted with three different procedures or types of equipment. One procedure is to collect that part of the plant that bears the seed. weigh it directly. or carry it to a stationary machine for threshing. The plant part may be removed by hand or may be collected with a machine. such as a mower with a collection basket mounted behind the sickle. The harvested sample may be threshed immediately or dried for a period of time before threshing. One popular type of stationary machine is the Vogel thresher. The plants pass vertically through the machine as they are threshed. For a second type of stationary thresher. the material passes through the threshing cylinder and falls on a sieve that helps separate the seed from the plant debris. Air is used to separate the seed and the plant debris in both types of machine.

The second procedure for harvesting plots is to use a self-propelled thresher specifically designed for small plots. The plant part with the seed is gathered into the machine and passes through a threshing cylinder. then the seed and plant debris are separated by sieves and air. The seed may be placed into a bag and saved or may be weighed immediately and discarded. Seed harvested from self-propelled machines generally is more subject to mixtures than that harvested with a stationary thresher.

The third type of equipment is a commercial combine modified for the harvest of small plots. A commercial unit is used only when the amount of seed harvested

FIELD-PLOT TECHNIQUES   285

from a plot is relatively large and is not saved for planting. Modifications of the commercial combine include reduction of the number of rows harvested and the addition of equipment for weighing the seed.

Data Collection

Usually a number of characters are measured on each plot, such as height, standability, and yield. The data may be recorded in a field book, then manually entered into the computer for statistical analysis. Alternatively, the information may be recorded in an electronic data collector and transferred directly to the computer. This saves time and reduces the possibility of human error. Plot and entry designations also can be recorded on labels that can be read into the data collector by an electronic scanner.

Data Analysis

Computers facilitate the selection of lines by summarizing data in whatever manner is beneficial to the breeder. They save an extensive amount of time, minimize human error, and permit data to be summarized in a short period of time.

REFERENCES

Baker, R. J., and R. I. H. McKenzie. 1967. Use of control plots in yield trials. *Crop Sci.* 7:335–337.

Burton, G. W. 1974. Recurrent restricted phenotypic selection increases forage yields of Pensacola bahiagrass. *Crop Sci.* 14:831–835.

Burton, J. W., and C. A. Brim. 1981. Recurrent selection in soybeans. III. Selection for increased percent oil in seeds. *Crop Sci.* 21:31–34.

Cochran, W. G. 1937. A catalogue of uniformity trial data. *Suppl. J. Roy. Stat. Soc.* 4:233–253.

Cochran, W. G., and G. M. Cox. 1957. Experimental design. 2d ed. John Wiley, New York.

Fasoulas, A. 1973. *A new approach to breeding superior yielding varieties.* Pub. No. 3, Department of General Plant Breeding. Aristotelian University of Thessaloniki, Greece.

———. 1979. *The honeycomb field designs.* Pub. No. 9, Department of General Plant Breeding. Aristotelian University of Thessaloniki, Greece.

———. 1981. *Principles and methods of plant breeding.* Pub. No. 11. Department of General Plant Breeding. Aristotelian University of Thessaloniki, Greece.

286   WALTER R. FEHR

Fehr. W. R.. and A. H. Probst. 1971. Effect of seed source on soybean strain performance for two successive generations. *Crop Sci.* 11:865–867.

Frey. K. J. 1965. The utility of hill plots in oat research. *Euphytica* 14:196–208.

Gardner. C. O. 1961. An evaluation of the effects of mass selection and seed irradiation with thermal neutrons on yield of corn. *Crop Sci.* 1:241–245.

Garland. M. L.. and W. R. Fehr. 1981. Selection for agronomic characters in hill and row plots of soybeans. *Crop Sci.* 21:591–595.

Gedge. D. L.. W. R. Fehr. and A. K. Walker. 1977. Intergenotypic competition between rows and within blends of soybeans. *Crop Sci.* 17:787–790.

Gomez, K. A.. and A. A. Gomez. 1984. *Statistical procedures for agricultural research.* 2d ed. John Wiley. New York.

LeClerg. E. L.. W. H. Leonard. and A. G. Clark. 1962. *Field plot technique.* 2d ed. Burgess Publishing. Minneapolis. Minn.

Mak. C.. B. L. Harvey. and J. D. Berdahl. 1978. An evaluation of control plots and moving means for error control in barley nurseries. *Crop Sci.* 18:870–873.

Peacock. H. A.. and B. S. Hawkins. 1970. Effect of seed source on seedling vigor. yield and lint characteristics of Upland cotton. *Gossypium hirsutum. L. Crop Sci.* 10:667–670.

Thorne. J. C.. and W. R. Fehr. 1970a. Effects of border row competition on strain performance and genetic variance in soybeans. *Crop Sci.* 10:605–606.

———. 1970b. Exotic germplasm for yield improvement in 2-way and 3-way soybean crosses. *Crop Sci.* 10:677–678.

Townley-Smith. T. F.. and E. A. Hurd. 1973. Use of moving means in wheat yield trials. *Can. J. Plt. Sci.* 53:447–450.

Wilcox. J. R. 1970. Response of soybeans to end-trimming at various growth stages. *Crop Sci.* 10:555–557.

814

1991; BOCKEI. et al. 1992). CHAKRABORTY et al. (1994) demonstrate that paternity can be determined in cases where the mother is unavailable for testing. LANG et al. (1993) partially reconstructed the DNA profile of a missing crocodile parent using profiles of the mother and progeny.

CHAKRABORTY et al. (1988) and SMOUSE and MEAGHER (1994) report that reliance upon exclusion alone has usually failed to unambiguously resolve paternity. Limitations have stemmed from the use of an insufficient number of independent hypervariable loci. Other statistical methods are therefore required to calculate the likelihood of paternity for each nonexcluded male (BERRY and GEISSER 1986; MEAGHER 1986; MEAGHER and THOMPSON 1986; THOMPSON and MEAGHER 1987; DEVLIN et al. 1988; BERRY 1991). MARSHALL et al. (1998) draw attention to the quality of data that is encountered practically in genotypic surveys. Maternal genetic data may or may not be available, data may be absent for some candidate males, data may be missing for some loci in some individuals, null alleles exist, and typing errors occur. Reconstructing or validating the pedigrees of varieties of cultivated plants often provides additional challenges because their phylogenies can reveal apparent exclusions that masquerade as non-Mendelian inheritance. For example, apparent exclusions can occur in circumstances where an individual is used as a parent prior to completion of the inbreeding process. The development of parent and progeny then continue on parallel but separate tracks thereby allowing the possibility that alleles that are subsequently lost through inbreeding in the parent can still become fixed in the progeny. It is also possible to create many offspring from a single mating and to use the same parent repeatedly in "backcrossing." Therefore, many individual inbred lines, varieties, or hybrids can be highly related. In consequence, there are numerous (and often very similar) pedigrees. The effective number of marker loci that can discriminate between alternate pedigrees is proportionally reduced as parents are increasingly related. Consequently, inbred lines can be more similar to one or more sister or other inbreds than those inbreds are to one or both of their parents.

It has not been usual to search among hundreds of individuals to identify the most probable maternal and paternal candidates for a specific progeny. Most studies of parentage are in circumstances where there is a priori information for at least one of the parents (usually the maternal parent). Limited availability of marker loci and the lack of very high-throughput genotyping systems offering inexpensive datapoint costs may have focused research on studies that involve relatively few individuals and where there is at least some a priori indication of parentage. Studies that have been conducted without a priori information on parentage include species where reproductive behavior renders identification of the maternal parent difficult or impossible. Examples include

those undertaken on birds that practice brood parasitism (ALDERSON et al. 1999) or extra-pair copulation (WETTON et al. 1992) or on species such as the wombat that are difficult to observe in the wild (TAYLOR et al. 1997).

Two circumstances favor a revised approach to the statistical analysis of pedigree. First, molecular marker technologies are rapidly developing and will allow numerous loci to be typed for thousands of individuals rapidly and inexpensively. A greater number and diversity of larger-scale studies of pedigree can be expected within the plant and animal kingdoms including individuals in which there is no prior knowledge of pedigree. A larger number of markers mean a greater chance for errors. Therefore, the second circumstance follows: Procedures that are efficient and robust in the face of apparent exclusions, missing data, and laboratory error are required.

The purpose of this article is to describe and evaluate a methodology that can be used to quantify the probability of parentage of hybrid genotypes. We focus on parentage because it is the primary focus of published literature and it is the easiest level of ancestry to understand. The method is robust in the face of mutation, pseudo-non-Mendelian inheritance (apparent exclusions) due to residual heterozygosity in parental seed sources, missing data, and laboratory error. The methodology has a number of advantages: (i) It can accommodate large datasets of possible ancestors (hundreds of inbreds or hybrids each profiled by >100 marker loci), (ii) it does not require prior knowledge about either parent of the hybrid of interest, (iii) it does not require independence of the markers, and (iv) it can successfully discriminate between many highly related and genetically similar genotypes. We demonstrate the effectiveness of this approach to identify inbred parents of maize (Zea mays L.) hybrids using simple sequence repeat (SSR) marker profiles for 54 maize hybrids together with their parental and grandparental genotypes included among a total of 586 inbred lines. The methodology is applicable to the investigation of parentage for all progeny developed from parental mating without subsequent generations of inbreeding.

MATERIALS AND METHODS

Algorithm: Consider an index hybrid whose parentage is unknown or in dispute. Inbreds in an available database are possible ancestors of the hybrid. The objective is to find the possible ancestors of the hybrid on the basis of probabilities of closest ancestry for each inbred on the basis of information from SSRs from the index hybrid and the inbreds. There is no reason to trim the database by removing inbreds thought to be unrelated to the index hybrid because their lack of relationship will be discovered.

Consider a pair of possible ancestors. inbred i and inbred j. There is nothing special about this particular pair as all pairs will be treated similarly. The process involves calculating the probability that inbreds i and j are in the hybrid's ancestry, repeating this for all pairs of inbreds in the database.

The basis of the algorithm is Bayes' rule (e.g., BERRY 1991, 1996). Let $P(i, j|SSRs)$ stand for the (posterior) probability that i and j are ancestors of the index hybrid given the information from the various SSRs. Let $P(i, j)$ stand for the unconditional (or prior) probability of the same event. Finally, $P(SSRs|i, j)$ is the probability of observing the various SSR results if in fact i and j are ancestors. Bayes' rule says

$$P(i, j|SSRs) = P(SSRs|i, j) \times P(i, j)/\sum[P(SSRs|u, v) \times P(u, v)],$$

where the sum in the denominator is over all pairs of inbreds, indexed by u and v. $P(SSRs|i, j) \times P(i, j)$ is one of the terms in the denominator. (To compute the denominator in the above expression, fix a particular order to the inbreds in the database and take $u < v$ in expressions involving the pair (u, v). If there are 586 inbreds, for example, then the number of pairs and the number of terms in the denominator is $586(587)/2 = 171,991$.) Inbreds i and j may be parents or grandparents or other types of relations or bear no relationship at all to the hybrid. If there are more than two ancestors in the database, such as both parents and all four grandparents, then the possible pairs involving these ancestors will generally have the highest posterior probabilities. If the hybrid's true parents are in the database, then as a pair they will typically have the highest overall posterior probability. If both i and j happen to be related to one particular parent of the hybrid, then as a pair their posterior probability will be low because they will not usually account for many of the alleles that are contributed by the other parent of the hybrid.

We will make the "no-prior-information" assumption that $P(u, v)$ is the same for all pairs (u, v). This implies that this factor is cancelled from both numerator and denominator in the above expression, giving:

$$P(i, j|SSRs) = P(SSRs|i, j)/\sum P(SSRs|u, v).$$

The problem is then to calculate a typical $P(SSRs|i, j)$. Assume inbreds i and j are both ancestors. We calculate the probability of observing the resulting hybrid under this assumption. We make no assumptions about relationships among the various inbreds. Other possible ancestors will be considered implicitly in the calculation by allowing their alleles to be introduced through breedings with i and j. However, the nature of such breedings is not specified. Suppose inbred i's alleles are (a, b). Each descendant of inbred i receives one of these two alleles or not. An immediate descendant receives one with probability 1 (barring mutations). A second generation descendant receives one of them with probability 0.5. And so on. Since degree of ancestry (if any) is unknown, we label the actual probability of passing on one of these alleles to be P. Similarly, an allele from inbred j has been passed down to the hybrid or not, and the probability of the former is P. In the following, P will be taken to equal 0.50, although we will also consider $P = 0.99$ in some of the calculations.

Assuming $P = 0.50$ is consistent with the closest ancestors in the database being grandparents. However, we are not interested in grandparents per se. If the closest ancestors in the database were parents, then as indicated above P should equal 1 (ignoring mutations and laboratory errors). Our primary concern is when the parents are not in the database. In this case P is no greater than 0.50. Assuming $P = 0.50$ is robust over the middle range of possible values of P. One way in which it is robust is if there may be mutations and laboratory errors, in which case P would have to be <1. Taking P to equal 0.50 levies little penalty against a particular pair in which there is an apparent exclusion from direct parentage. Therefore taking P to be <1 means that if the true parents are in the database then they will not be ruled out if there happen to be mutations and laboratory errors. And if the closest ancestors in the database are more remote than grandparents, they

are likely to be identified because they will usually have the fewest mismatches of the lines considered.

When i and j are ancestors there are four possibilities: (1) The alleles of both inbreds i and j were passed to the hybrid, (2) inbred i came through but not inbred j, (3) inbred j came through but not inbred i, and (4) neither inbred came through. Assuming independence, these have respective probabilities P^2, $P(1 - P)$, $P(1 - P)$, $(1 - P)^2$. In the case $P = 0.50$, all of these probabilities equal 0.25.

An instance of the law of total probability (Sec. 5.3, BERRY 1996) is that the probability of observing a hybrid's alleles is the average of the conditional probability of this event given the above four cases. The simplest of the four cases is the first possibility: Assuming the hybrid's alleles are passed down directly from both inbreds, the probability of observing the hybrid's genotype is either 1 or 0 depending on whether the hybrid shares both inbreds' alleles. (It is especially easy when both inbreds are homozygous.) The other three cases require an assumption regarding the possibility that an inbred's allele is not passed to the hybrid but is interrupted by a mutation, a laboratory error, or intervening breeding. We regard such an allele as being selected from all known alleles with probability 1/(number of alleles), where the number of alleles is the total number of alleles known to exist at the locus in question. An alternative approach would be to use the allelic proportions that are present in the database (or in another database). However, the lines in the database may not be randomly selected from any population. For example, a line that has been highly used in breeding would have many derivative lines in the database, in which case the frequencies of its alleles will be artificially inflated. Assuming equal probabilities for the various alleles at a given locus is robust in the sense that it is not affected by adding and dropping lines from the database.

There are many cases to consider when computing the probability of observing a hybrid's alleles, depending on the zygosity of the hybrid and the inbreds, and allowing for the possibility of missing alleles or "extra alleles" in the assessment of the hybrid and inbred genotypes. These possibilities are too numerous to list. Instead we give three simple examples. All the examples have homozygous inbreds, the most common case. And each of the three hybrids has two alleles, again the most common case. We suppose that the measured alleles for three SSRs and a particular trio of hybrid and ancestor inbreds are as we have indicated in Table 1.

For SSR 1 there are three known alleles, one in addition to alleles a and b that are listed for the three lines (hybrid, inbred i, and inbred j) in Table 1. For SSR 2 and SSR 3 there are two known alleles in addition to those listed. The calculations in the right half of Table 1 will now be explained. Implicit in calculating $P(SSR|i, j)$ is the assumption—required in both the numerator and denominator of Bayes' rule—that inbreds i and j are ancestors of the hybrid. Consider SSR 1. In case 1 above, both ancestors' alleles (as measured by the laboratory process) are assumed to pass to the index hybrid, and so in this case the hybrid is necessarily ab. The probability of observing the actual hybrid's genotype is 1 for case 1, as shown in Table 1. In case 2, we assume that inbred i's allele passes to the hybrid but inbred j's does not. Indeed, the hybrid has an a allele. The probability of observing a b as the other allele is 1/(number of alleles) = 1/3, as shown in Table 1. Case 3 is similar. In case 4, neither ancestor allele is passed to the hybrid: the probability of observing the hybrid's genotype (or any heterozygous genotype) is $2(1/3)(1/3) = 2/9$. Since $P = 0.50$, the overall (unconditional) probability in the rightmost column (17/36) is the simple average of the four cases, as indicated in Table 1.

For SSR 2 and SSR 3 the calculations are similar. For SSR 2 there is some evidence against pair (i, j) being ancestors,

TABLE 1

Probability of observing a hybrid's alleles using three sample SSRs and four possible combinations (cases) of alleles passed, assuming that inbreds i and j are ancestors of the hybrid

SSR	No. of alleles	Hybrid	Inbred i	Inbred j	Probability of observing the hybrid's genotype				Overall probability $P(SSR\|i,j)$
					Case 1 i, j	Case 2 $i, not j$	Case 3 $not i, j$	Case 4 $not i, not j$	
1	3	ab	aa	Bb	1	1/3	1/3	2/9	17/36
2	5	bd	bb	Cr	0	1/5	0	2/25	7/100
3	6	ab	cc	Dd	0	0	0	2/36	2/144

SSR, simple sequence repeat marker profile.

but it is not conclusive. For SSR 3 there is even less evidence favoring pair (i, j). It would not take many SSRs with evidence similar to that for SSR 3 to essentially rule out this pair—provided that other pairs are not similarly inconsistent.

To find the overall $P(SSRs|i, j)$, multiply the individual $P(SSR|i, j)$ over the various SSRs. There are purely computational issues to address. Each $P(SSR|i, j)$ is a number between 0 and 1. When there are a great many SSRs, the product of these numbers will be vanishingly small. To lessen problems with computational underflow, for each SSR we multiply $P(SSR|u, v)$ by the same constant for each pair (u, v): the inverse of the largest possible such probability. For example, since 17/36 is the largest probability for a heterozygous hybrid at an SSR having three alleles (as is the case for SSR 1 in Table 1), we multiply all factors $P(SSR1|u, v)$ by 36/17. To eliminate remaining problems with underflow, we do calculations using logarithms (adding instead of multiplying) and take antilogs at the end.

The probability $P(SSR|u, v)$ is calculated for all (u, v) pairs and summed over all possible pairings in the database, including that for the inbred pair under consideration: (i, j). This gives the denominator in the expression for $P(i, j|SSRs)$.

To determine the probability that any particular inbred, say inbred i, is the closest ancestor of the index hybrid, sum $P(SSR|i, v)$ over all inbreds v with $v =/ i$. Call this $P(i|SSRs)$. The maximum of $P(i|SSRs)$ for any inbred i is 1. But since there is one closest ancestor on each side of the family, the sum of $P(i|SSRs)$ over all inbreds i is 2. If there is a particular pair (i, j) for which $P(i, j|SSRs)$ is close to 1 then both $P(i|SSRs)$ and $P(j|SSRs)$ separately will be close to 1.

SSR data: DNA was extracted from 54 maize hybrids and from 386 maize inbreds. All of the hybrids and most inbreds are proprietary products of Pioneer Hi-Bred International; some important publicly bred inbred lines were also included. The inbred parents and grandparents of each hybrid were included within the set of inbreds. Other inbreds that were genotyped include many that are highly related by pedigree to parents and grandparents of the hybrids. The hybrids were chosen because each has a pedigree that is known to us and collectively they represent a broad array of diversity of maize germplasm that is currently grown in the United States ranging from early to late maturity.

A total of 195 SSR loci were used in this study following procedures described in SMITH et al. (1997), but modified as described below. SSR loci were chosen on the basis that they individually have been shown to have a high power of discrimination among maize inbred lines and collectively they provide for a sampling of diversity for each chromosome arm. Of these SSR loci, the following numbers (in parentheses) were located on individual maize chromosomes as follows: 1 (35), 2 (26), 3 (22), 4 (20), 5 (16), 6 (9), 7 (6), 8 (18), 9 (12), and 10

(14); 17 SSR loci have not yet been mapped. The correlations among the loci are unknown and are irrelevant for our methodology.

Sequence data for primers that allow many of these (and other) SSR loci to be assayed are available at website http://www.agron.missouri.edu. All primers were designed to anneal and amplify under a single set of conditions for PCR in 10-µl reactions. Genomic DNA (10 ng) was amplified in 1.5 mM reactions. 50 mM KCl, 10 mM Tris-Cl (pH 8.3) using 0.3 units AmpliTaq Gold DNA polymerase (PE Corporation) oligonucleotide primer pairs (one primer of each pair was fluorescently labeled) at 0.17 µM and 0.2 mM dNTPs. This mixture was incubated at 95° for 10 min (hot start); amplified using 45 cycles of denaturation at 95° for 50 sec, annealing at 60° for 50 sec, extension at 72° for 85 sec; and then terminated at 72° for 10 min. A water bath thermocycler manufactured at Pioneer Hi-Bred International was used for PCR reactions. PCR products were prepared for electrophoresis by diluting 3 µl of each product to a total of 27 µl using a combination of PCR products generated from other loci for that same maize genotype (multiplexing) and/or dH20. Dilution of 1.5 µl of this mixture to 5 µl with gel loading dye was performed; it was then electrophoresed at 1700 V for 1.5 hr on an ABI model 377 automated DNA sequencer equipped with GENE-SCAN software v. 3.0 (PE-Applied Biosystems, Foster City, CA).

PCR products were sized automatically using the "local Southern"-sizing-algorithm (ELDER and SOUTHERN 1987). After sizing of PCR products using GeneScan, alleles were assigned using Genotyper software (PE-Applied Biosystems). Generally, allele assignations for each locus were made on the basis of histogram plots consisting of 0.5-bp bins. Breaks between the histogram plots of >1 bp were generally considered to constitute separation between allele bins; however, other criteria, such as the presence of the nontemplate-directed addition of adenine (+A addition) and naturally occurring 1-bp alleles, were used on a marker-by-marker basis to define the allele dictionary. All allele scores were made without knowing the identities of the maize genotypes.

RESULTS

Table 2 presents the probability of closest ancestry of the top five ranking inbred lines for each of 5 hybrids at $P = 0.50$ (Table 2A) and $P = 0.99$ (Table 2B). Probabilities of ancestry are shown for all 54 hybrids and the top ranking inbreds in Figure 1: $P = 0.50$ (Figure 1a) and $P = 0.99$ (Figure 1b). Results for the hybrids presented in Table 2 are featured at the top of Figure 1.

TABLE 2

Probability of ancestry of five hybrids using data obtained from 50, 100, and 195 SSR loci

Hybd.	50 loci			100 loci			195 loci		
	Inbd.	Prob.	SE	Inbd.	Prob.	SE	Inbd.	Prob.	SE
A. Assuming P = 0.50									
3417	SP1	0.9607	0.0125	P1	0.8749	0.0252	P1	1.0000	E-07
	P2	0.8077	0.1965	P2	0.8141	0.2235	P2	0.9937	0.0033
	D2P2	0.1016	0.1038	D1P2	0.1859	0.2235	D1P2	0.0043	0.0033
	D1P2	0.0907	0.0927	SP1	0.1243	0.025	D2P2	E-06	E-06
	P1	0.032	0.0125	D1P1	0.0009	0.0002	SP1	E-06	E-07
3525	P1	0.8345	E-07	P1	0.9999	<E-20	P1	1.0000	<E-20
	P2	0.8188	E-07	P2	0.5437	<E-20	P2	0.9635	0.0528
	D1P2	0.1699	E-07	D1P2	0.4563	<E-20	D1P2	0.0365	0.0528
	GP1	0.1441	E-07	GP1	E-07	E-18	SP1	E-15	<E-20
	GP2	0.0110	E-08	SP1	E-07	<E-20	GP2	E-16	<E-20
3556	P1	1.0000	E-06	P1	0.9999	E-10	P1	1.0000	<E-20
	P2	0.9616	E-08	P2	0.9997	E-10	P2	1.0000	<E-20
	D1P2	0.0340	E-10	D1P2	0.0003	E-14	D1P2	E-09	<E-20
	GP2	0.0043	E-09	D2P2	E-05	E-15	D2P2	E-14	<E-20
	D2P2	0.0002	E-10	D3P2	E-06	E-17	GGP2	E-17	E-17
3905	D1P1	0.9822	E-08	D1P1	0.9803	0.0058	P1	1.0000	E-08
	SP2	0.4927	E-07	SP2	0.6280	0.0976	D1P2	1.0000	E-06
	D2P2	0.2336	E-07	D1P2	0.2321	0.0617	D2P2	E-06	E-06
	D1P2	0.1622	E-07	D2P2	0.1317	0.0372	P2	E-07	E-13
	P2	0.0565	E-07	P1	0.0197	0.0058	D3P2	E-10	E-16
3940	P2	0.9997	0.0001	P2	0.9999	E-05	P2	1.0000	E-09
	D1P2	0.9203	0.0009	P1	0.9970	0.0011	P1	1.0000	E-09
	P1	0.0648	E-05	D1P2	0.0030	0.0011	D1P2	E-11	E-11
	D1P1	0.0127	E-05	D2P2	0.0001	E-05	DP1P2	E-17	E-17
	DP1P2	0.0014	0.0009	DP1P2	0.0001	E-07	D2P2	E-19	E-18
B. Assuming P = 0.99									
3417	SP1	0.9995	0.0001	P1	0.9999	E-05	P1	0.9999	E-08
	P2	0.8836	0.1658	P2	0.9938	0.0107	P2	0.9999	E-08
	D1P2	0.0722	0.1029	D1P2	0.0061	0.0107	D1P2	E-11	E-11
	D2P2	0.0441	0.0628	D1P1	E-05	E-06	D2P2	E-14	E-14
	P1	0.0004	0.0001	SP1	E-05	0	SP1	E-20	E-21
3525	P1	0.9999	0	P1	0.9999	0	P1	1.0000	0
	P2	0.8991	0	D1P2	0.9749	0	P2	0.6135	0.4446
	D1P2	0.1008	E-11	P2	0.025	0	D1P2	0.3864	0.4446
	GP1	E-05	0	D2P2	E-20	0	GP2	E-48	0
	GP2	E-06	E-17	SP1	E-24	0	D2P2	E-49	0
3556	P1	1.0000	0	P1	1.0000	0	P1	0.9999	0
	P2	0.9996	0	P2	0.9999	0	P2	0.9999	0
	D1P2	0.0003	0	D1P2	E-09	0	D1P2	E-22	0
	D1P1	E-11	0	D3P1	E-21	0	D2P1	E-49	0
	D2P1	E-13	0	D2P1	E-21	0	D3P1	E-54	0
3905	D1P1	0.9999	0	D1P1	0.9999	E-08	P1	1.0000	E-09
	P2	0.9992	0	P2	0.9999	E-06	P2	0.9947	E-09
	SP2	0.0006	0	D1P2	E-06	E-06	D1P2	0.0052	E-11
	D1P2	E-05	0	SP2	E-07	E-13	D2P2	E-18	E-18
	D2P2	E-06	0	D2P2	E-09	E-10	D1P1	E-25	E-25
3940	P2	0.9999	E-08	P2	1.0000	E-08	P1	1.0000	E-09
	D1P2	0.9999	E-08	P1	0.9999	E-05	P2	1.0000	E-09
	P1	E-06	E-13	D1P2	E-05	E-05	D1P2	E-24	E-24
	D1P1	E-08	E-15	D2P2	E-12	E-11	DP1P2	E-44	E-44
	DP1P2	E-12	E-12	DP1P2	E-21	E-21	D2P2	E-50	E-49

Hvbd., hybrid; Inbd., inbred; Prob., probability; SE, standard error, referring to the variability in the results of the runs; P1, parent one; P2, parent two; SP1/SP2, full sibling of parent one/parent two; D/P1/D/P2, derivatives of parent one/parent two, index *i* for distinct inbred lines; DP1P2, derivatives of both parent one and parent two.

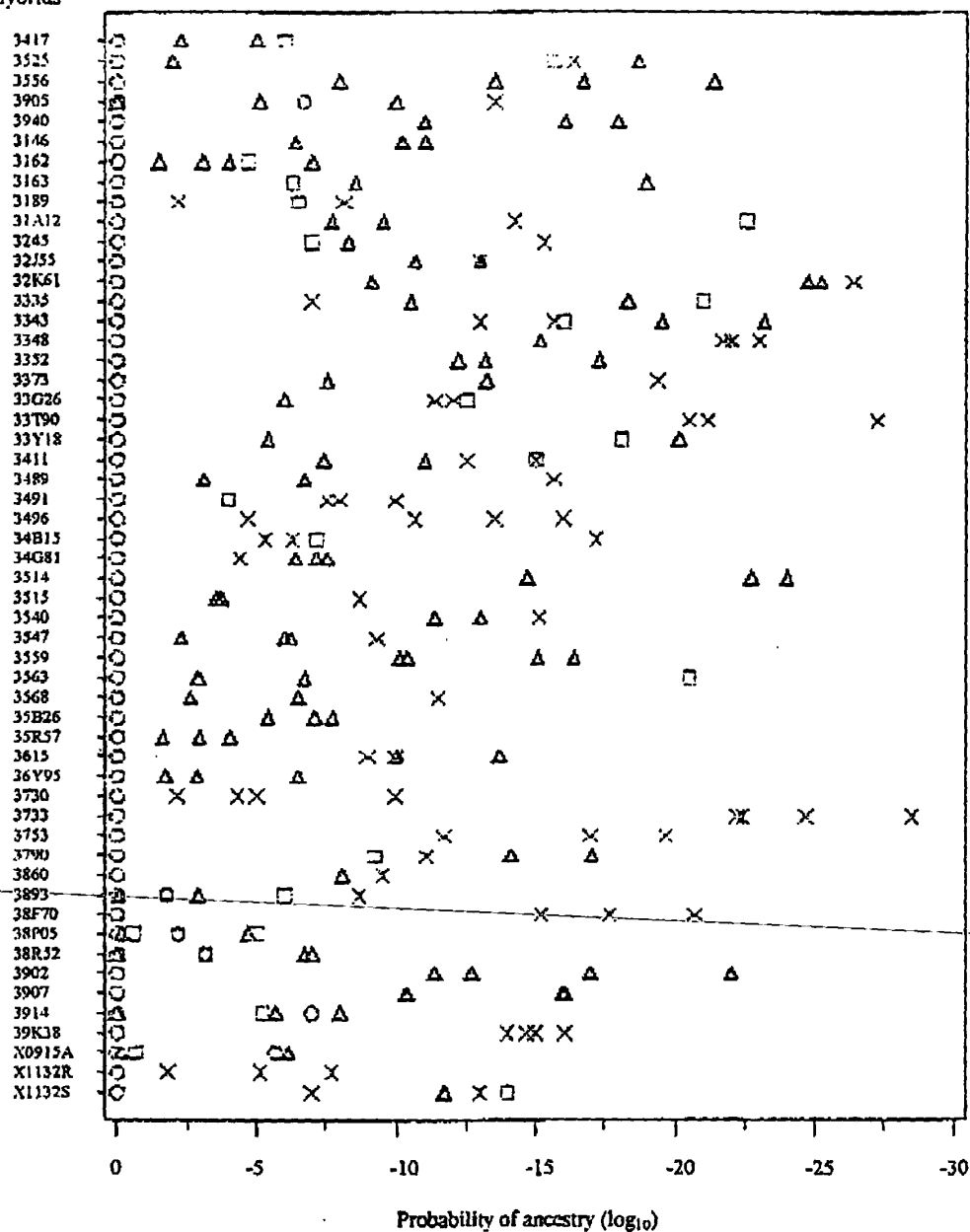FIGURE 1.—(a) Probabilities of ancestry, assuming $P = 0.50$, for all 54 hybrids and top ranking inbreds—those with probability of ancestry at least 10^{-30}. (b) Probabilities of ancestry, assuming $P = 0.99$, for all 54 hybrids and top ranking inbreds—those with probability of ancestry at least 10^{-30}.
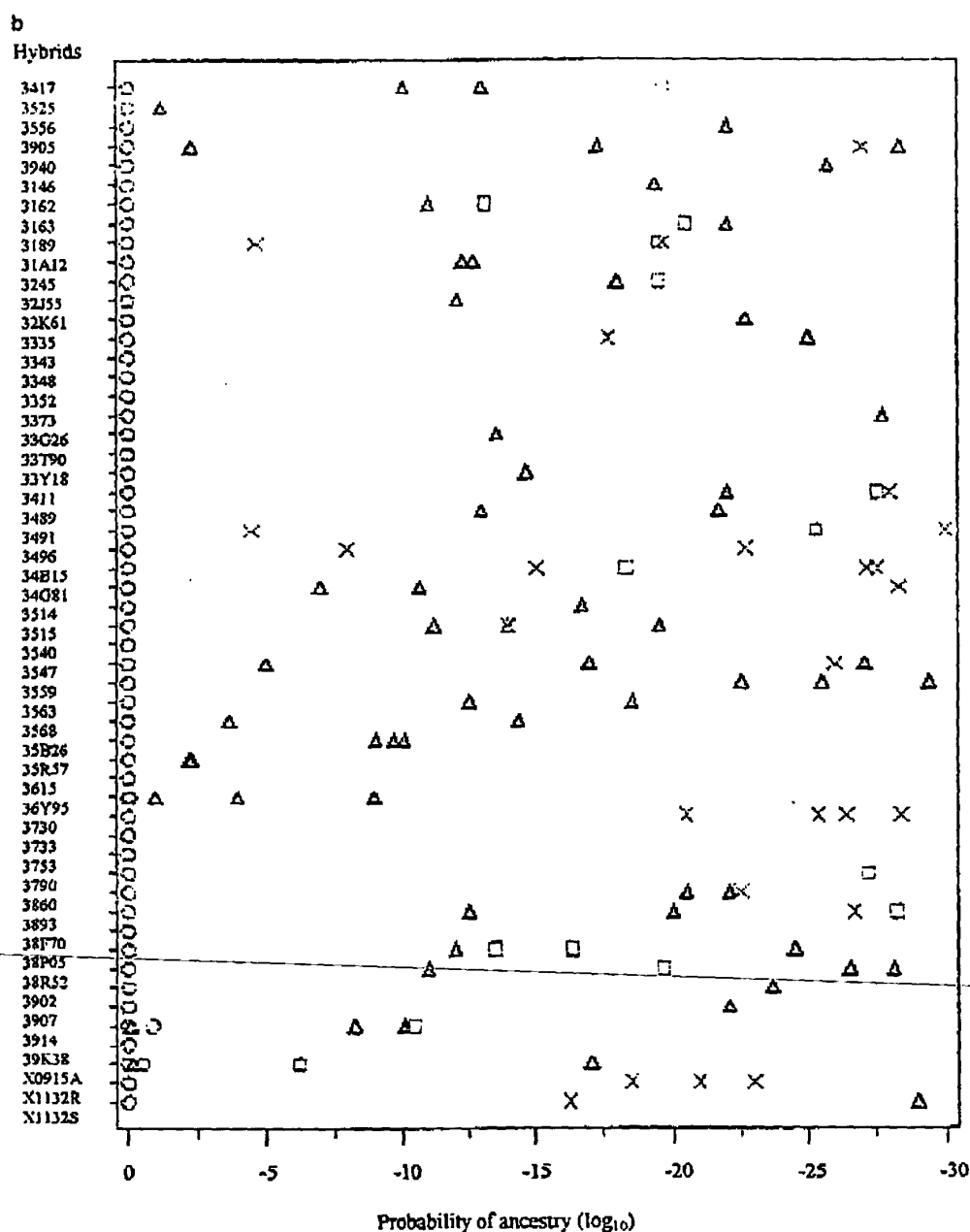
FIGURE 1.—Continued.

When the algorithm used $P = 0.50$, the two correct parents were identified as highest in probability for 48 (89%) hybrids (Figure 1). For each of 6 hybrids (3893, 38P05, 38R52, 3905, 3914, and X0915A), one parent ranked in the top two places. The other parent was supplanted either by a sister inbred or by an inbred that was a direct progeny of that parent. Overall, 102 (94%) of 108 parental inbreds were correctly identified. For hybrids where both parents ranked first or second, the range of probabilities for parental lines that ranked first from among all other inbreds ranged from 1.0000 to 0.9997; parental lines ranking second ranged from

820

1.0000 to 0.9653. For 35 hybrids, both parents had probabilities of ancestry in excess of 0.999. Probabilities of ancestry for nonparents that ranked in first or second places were from 0.9999 to 0.7054. For the majority of hybrids, the probability of the third and highest ranked nonparental inbred was at or below E-06. This indicates that there is usually very little uncertainty about closest ancestors.

When the algorithm used $P = 0.99$ to examine each of the 54 hybrids, both parents were correctly identified for 52 (96%) of hybrids and for 98% (102/104) of the parents across all hybrids (Figure 1). Two hybrids (3914 and X0915A), in which one parent was not ranked in the top two, were also in the subset not ranked in the top two assuming $P = 0.30$ (above). In both cases their ranks improved (both to third rank) and the actual parent was supplanted by an inbred that was a direct progeny of the corresponding parental line. For 49 hybrids, both parents had probabilities of ancestry in excess of 0.999. Among the 5 hybrids having a parent ranking second with a probability of ancestry below 0.999, the lowest of these probabilities was 0.8976 and the highest probability for a third ranking nonparent was 0.1023. For most hybrids the probability for the third and highest ranked nonparental inbred was at or below E-10.

Table 2 also addresses data analysis in circumstances where heterozygous loci occur in inbred lines or where a hybrid is scored for the presence of more than two alleles per locus. The presence of more than a single allele per locus in inbred lines is an infrequent occurrence in well-maintained inbred development and seed increase programs but is possible because ~3–5% of loci can still be segregating and unintended pollination from genotypes not designated as parents of the hybrid can occur. For hybrids, more than two alleles per locus can be scored when DNA is extracted from a bulk of individual plants and because inbred parents are not homozygous due either to residual heterozygosity or to contamination or because one or more direct parents of the hybrid are themselves hybrids. The presence of more than one allele per locus in an inbred line and more than two alleles per locus in a hybrid therefore can be accommodated by multiple runs of the algorithm, each with a random choice of two alleles per locus. Consequently, standard errors in the case of analyzing data from 195 loci tend to be very small because there were few loci where an inbred or hybrid sample (from a bulk of individual plants) was scored for more than two alleles.

MARSHALL *et al.* (1998) have drawn attention to errors that can be encountered in genotyping surveys. These errors include missing data, null alleles, and typing errors. We therefore investigated the robustness of the algorithm by examining the effects of modifications in the data for five hybrids (3417, 3525, 3556, 3905, and

3940). First, we reduced the number of SSRs used, from the full set of 195 to 100 and then to 50 (Table 2). Use of 50 loci generated incorrect rankings of one parent for each of two hybrids (3417 and 3940) and for both parents of one hybrid (3905). All of these most highly ranked nonparental inbreds were closely related to the true parents for each of the respective hybrids; six different inbred lines were involved. Four were direct progeny of the true parents (one with additional backcrosses from the true parent) and two were full sisters (from a cross of highly related inbreds) of the actual parent of the hybrid. Using 100 loci resulted in correct parental rankings for all hybrids except for 3905 where neither parent ranked in first or second place. Four inbreds outranked the true parents of 3905. All four nonparents were closely related to the respective true parents; three were direct progeny of the true parent of the hybrid (one with additional backcrossing to that parent) and one was a full sister of the true parent. Use of data from all 195 loci corrected the placement for one of the parents of hybrid 3905. Two inbreds that were not parents of this hybrid remained ranked more highly than one of the true parents. Both were direct progeny of one of the true parents, and one of these inbreds had additional backcrossing to that parent in its pedigree.

To address the consequences of laboratory and other sources of error, we artificially compromised data quality beyond the level originally provided by eliminating specific proportions of alleles that had been scored (establishing scenarios where various numbers of SSR alleles were not scored) and by misscoring other alleles (establishing scenarios where various numbers of SSR alleles were scored incorrectly). We also combined the scenarios of missing data and wrongly scored data. Table 3 contains a summary of the results of making these modifications in the data. For all modifications we used data from all SSR loci and we also randomly chose SSR loci to create subsets of 50 and 100 loci. In each case, the program was run 20 times for each hybrid/set of loci. When all 195 loci were examined, replications differed only according to the particular choice of alleles for loci where more than two alleles had been scored.

To evaluate robustness in the face of missing data or mistyped data, we simulated individual and combined categories of these data in the hybrid and all inbred lines at levels of 2, 5, 10, and 25% of the alleles for each of five hybrids and all inbreds beyond the level of error as originally scored by the laboratory. We examined the effects of these levels and types of error for three sizes of database: 50 loci, 100 loci, and all 195 scored loci. The same five hybrids considered in Table 2 were investigated: 3417, 3525, 3556, 3905, and 3940. One of these hybrids (3905) was chosen because one of its parents did not rank among the top two places even when the complete and unmodified data from all SSR loci were used.

Examples of robustness in the face of additional error

Assessing Probability of Ancestry Using Simple Sequence Repeat Profiles: Applications to Maize Hybrids and Inbreds

Donald A. Berry,*,1 Jon D. Seltzer,† Chongqing Xie,‡ Deanne L. Wright‡ and J. Stephen C. Smith‡

*The University of Texas M. D. Anderson Cancer Center, Houston, Texas 77030, †Third Wave Technologies, Inc., Madison, Wisconsin
53719 and ‡Pioneer Hi-Bred International, Inc., Johnston, Iowa 50131

ABSTRACT

Determination of parentage is fundamental to the study of biology and to applications such as the identification of pedigrees. Limitations to studies of parentage have stemmed from the use of an insufficient number of hypervariable loci and mismatches of alleles that can be caused by mutation or by laboratory error and that can generate false exclusions. Furthermore, most studies of parentage have been limited to comparisons of small numbers of specific parent-progeny triplets thereby precluding large-scale surveys of candidates where there may be no prior knowledge of parentage. We present an algorithm that can determine the probability of parentage in circumstances where there is no prior knowledge of pedigree and that is robust in the face of missing data or mistyped data. We present data from 54 maize hybrids and 586 maize inbreds that were profiled using 195 SSR loci including simulations of additional levels of missing and mistyped data to demonstrate the utility and flexibility of this algorithm.

DETERMINATION of parentage is fundamental to the study of reproductive and behavioral biology. The increasing availability of highly discriminant genetic markers for many diverse species provides the potential to uniquely characterize individuals at numerous loci and to unambiguously resolve parentage where genealogical relationships are unknown, in error, or in dispute.

Identification of parent-progeny relationships in wild populations of animals and plants provides insights into the success of various reproductive strategies (ELLSTRAND 1984; SMOUSE and MEAGHER 1994; ALDERSON et al. 1999) and has allowed for the implementation of management programs to conserve genetic diversity (MILLER 1975; RANNALA and MOUNTAIN 1997). The association of pedigree with physical appearance or performance in domesticated animals and plants allows parents that have contributed favorable alleles for desirable traits through selective breeding programs to be identified (BOWERS and MEREDITH 1997; SEFC et al. 1998; VANKAN and FADDY 1999). These applications of associative genetics facilitate further progress in genetic improvement through breeding. Establishment of parentage is also useful to secure legal rights of guardianship in humans, to help protect intellectual property in plant varieties, to validate breed pedigrees of domesticated animals, to protect stocks of fish, and to identify provenance of meat that is available in supermarkets

1 Corresponding author: Department of Biostatistics, The University of Texas M. D. Anderson Cancer Center, 1515 Holcombe Blvd., Box 447, Houston, TX 77030-4009. E-mail: dberry@mdanderson.org

(GOTZ and THALLER 1998; PRIMMER et al. 2000; WHITE et al. 2000).

Most studies of pedigree have utilized exclusion analysis where the molecular marker genotypes of either one or a restricted number of potential triplets of offspring and putative parents are compared. Often the identity of the mother is not in question; the maternal profile is subtracted from that of the offspring and the deduced paternal profile is then compared with candidate father genotypes (ELLSTRAND 1984; HAMRICK and SCHNABEL 1985). Individuals who could not have contributed the paternal genotype are excluded; the remainder are possible parents. Nonpaternity in humans is generally declared only on the basis of exclusions exhibited by at least two unlinked and independent loci. This criterion of exclusion reduces the likelihood of a false declaration of nonpaternity on the basis of marker results that are actually due to mutation within the phylogeny. BERRY et al. (1998) show that evidence of nonpaternity should require exclusions at loci on different chromosomes to avoid erroneous conclusions that would be made due to nondisjunction at meiosis leading to uniparental inheritance. A requirement for at least three independent exclusions to declare nonpaternity in humans has also been instituted (GUNN et al. 1997). In studies of natural populations of animals or plants where numerous parent-progeny triplets are examined it is usual to accept a single exclusionary event as evidence of nonpaternity (MARSHALL et al. 1998). Paternity testing has been extended to situations where DNA from either parent is unavailable. For example, paternity can still be established in circumstances where the putative father is deceased but his parents are still alive (HELMINEN et al.

Attachment B
Serial No. 09/759,800

811

1991; Bocker. et al. 1992; Chakraborty et al. (1991) demonstrate that paternity can be determined in cases where the mother is unavailable for testing. Lang et al. (1993) partially reconstructed the DNA profile of a missing crocodile parent using profiles of the mother and progeny.

Chakraborty et al. (1988) and Smouse and Meagher (1994) report that reliance upon exclusion alone has usually failed to unambiguously resolve paternity. Limitations have stemmed from the use of an insufficient number of independent hypervariable loci. Other statistical methods are therefore required to calculate the likelihood of paternity for each nonexcluded male (Berry and Geisser 1986; Meagher 1986; Meagher and Thompson 1986; Thompson and Meagher 1987; Devlin et al. 1988; Berry 1991). Marshall et al. (1998) draw attention to the quality of data that is encountered practically in genotypic surveys. Maternal genetic data may or may not be available, data may be absent for some candidate males, data may be missing for some loci in some individuals, null alleles exist, and typing errors occur. Reconstructing or validating the pedigrees of varieties of cultivated plants often provides additional challenges because their phylogenies can reveal apparent exclusions that masquerade as non-Mendelian inheritance. For example, apparent exclusions can occur in circumstances where an individual is used as a parent prior to completion of the inbreeding process. The development of parent and progeny then continue on parallel but separate tracks thereby allowing the possibility that alleles that are subsequently lost through inbreeding in the parent can still become fixed in the progeny. It is also possible to create many offspring from a single mating and to use the same parent repeatedly in "backcrossing." Therefore, many individual inbred lines, varieties, or hybrids can be highly related. In consequence, there are numerous (and often very similar) pedigrees. The effective number of marker loci that can discriminate between alternate pedigrees is proportionally reduced as parents are increasingly related. Consequently, inbred lines can be more similar to one or more sister or other inbreds than those inbreds are to one or both of their parents.

It has not been usual to search among hundreds of individuals to identify the most probable maternal and paternal candidates for a specific progeny. Most studies of parentage are in circumstances where there is *a priori* information for at least one of the parents (usually the maternal parent). Limited availability of marker loci and the lack of very high-throughput genotyping systems offering inexpensive datapoint costs may have focused research on studies that involve relatively few individuals and where there is at least some *a priori* indication of parentage. Studies that have been conducted without *a priori* information on parentage include species where reproductive behavior renders identification of the maternal parent difficult or impossible. Examples include

those undertaken on birds that practice brood parasitism (Anderson et al. 1999) or extra-pair copulation (Wetton et al. 1992) or on species such as the wombat that are difficult to observe in the wild (Taylor et al. 1997).

Two circumstances favor a revised approach to the statistical analysis of pedigree. First, molecular marker technologies are rapidly developing and will allow numerous loci to be typed for thousands of individuals rapidly and inexpensively. A greater number and diversity of larger-scale studies of pedigree can be expected within the plant and animal kingdoms including individuals in which there is no prior knowledge of pedigree. A larger number of markers mean a greater chance for errors. Therefore, the second circumstance follows: Procedures that are efficient and robust in the face of apparent exclusions, missing data, and laboratory error are required.

The purpose of this article is to describe and evaluate a methodology that can be used to quantify the probability of parentage of hybrid genotypes. We focus on parentage because it is the primary focus of published literature and it is the easiest level of ancestry to understand. The method is robust in the face of mutation, pseudo-non-Mendelian inheritance (apparent exclusions) due to residual heterozygosity in parental seed sources, missing data, and laboratory error. The methodology has a number of advantages: (i) It can accommodate large datasets of possible ancestors (hundreds of inbreds or hybrids each profiled by >100 marker loci), (ii) it does not require prior knowledge about either parent of the hybrid of interest, (iii) it does not require independence of the markers, and (iv) it can successfully discriminate between many highly related and genetically similar genotypes. We demonstrate the effectiveness of this approach to identify inbred parents of maize (*Zea mays* L.) hybrids using simple sequence repeat (SSR) marker profiles for 54 maize hybrids together with their parental and grandparental genotypes included among a total of 586 inbred lines. The methodology is applicable to the investigation of parentage for all progeny developed from parental mating without subsequent generations of inbreeding.

MATERIALS AND METHODS

Algorithm: Consider an index hybrid whose parentage is unknown or in dispute. Inbreds in an available database are possible ancestors of the hybrid. The objective is to find the probabilities of closest ancestry for each inbred on the basis of information from SSRs from the index hybrid and the inbreds. There is no reason to trim the database by removing inbreds thought to be unrelated to the index hybrid because their lack of relationship will be discovered.

Consider a pair of possible ancestors, inbred *i* and inbred *j*. There is nothing special about this particular pair as all pairs will be treated similarly. The process involves calculating the probability that inbreds *i* and *j* are in the hybrid's ancestry, repeating this for all pairs of inbreds in the database.

The basis of the algorithm is Bayes' rule (e.g., Berry 1991, 1996). Let $P(i, j|SSRs)$ stand for the (posterior) probability that i and j are ancestors of the index hybrid given the information from the various SSRs. Let $P(i, j)$ stand for the unconditional (or prior) probability of the same event. Finally, $P(SSRs|i, j)$ is the probability of observing the various SSR results if in fact i and j are ancestors. Bayes' rule says

$$P(i, j|SSRs) = P(SSRs|i, j) \times P(i, j) \sum [P(SSRs|u, v) \times P(u, v)].$$

where the sum in the denominator is over all pairs of inbreds, indexed by u and v. $P(SSRs|i, j) \times P(i, j)$ is one of the terms in the denominator. (To compute the denominator in the above expression, fix a particular order to the inbreds in the database and take $u < v$ in expressions involving the pair (u, v). If there are 586 inbreds, for example, then the number of pairs and the number of terms in the denominator is $586(587)/2 = 171,991$.) Inbreds i and j may be parents or grandparents or other types of relations or bear no relationship at all to the hybrid. If there are more than two ancestors in the database, such as both parents and all four grandparents, then the possible pairs involving these ancestors will generally have the highest posterior probabilities. If the hybrid's true parents are in the database, then as a pair they will typically have the highest overall posterior probability. If both i and j happen to be related to one particular parent of the hybrid, then as a pair their posterior probability will be low because they will not usually account for many of the alleles that are contributed by the other parent of the hybrid.

We will make the "no-prior-information" assumption that $P(u, v)$ is the same for all pairs (u, v). This implies that this factor is cancelled from both numerator and denominator in the above expression, giving:

$$P(i, j|SSRs) = P(SSRs|i, j)/\sum P(SSRs|u, v).$$

The problem is then to calculate a typical $P(SSRs|i, j)$. Assume inbreds i and j are both ancestors. We calculate the probability of observing the resulting hybrid under this assumption. We make no assumptions about relationships among the various inbreds. Other possible ancestors will be considered implicitly in the calculation by allowing their alleles to be introduced through breedings with i and j. However, the nature of such breedings is not specified. Suppose inbred i's alleles are (a, b). Each descendant of inbred i receives one of these two alleles or not. An immediate descendant receives one with probability 1 (barring mutations). A second generation descendant receives one of them with probability 0.5. And so on. Since degree of ancestry (if any) is unknown, we label the actual probability of passing on one of these alleles to be P. Similarly, an allele from inbred j has been passed down to the hybrid or not, and the probability of the former is P. In the following, P will be taken to equal 0.50, although we will also consider $P = 0.99$ in some of the calculations.

Assuming $P = 0.50$ is consistent with the closest ancestors in the database being grandparents. However, we are not interested in grandparents per se. If the closest ancestors in the database were parents, then as indicated above P should equal 1 (ignoring mutations and laboratory errors). Our primary concern is when the parents are not in the database. In this case P is no greater than 0.50. Assuming $P = 0.50$ is robust over the middle range of possible values of P. One way in which it is robust is if there may be mutations and laboratory errors, in which case P would have to be <1. Taking P to equal 0.50 levies little penalty against a particular pair in which there is an apparent exclusion from direct parentage. Therefore taking P to be <1 means that if the true parents are in the database then they will not be ruled out if there happen to be mutations and laboratory errors. And if the closest ancestors in the database are more remote than grandparents, they

are likely to be identified because they will usually have the fewest mismatches of the lines considered.

When i and j are ancestors there are four possibilities: (1) The alleles of both inbreds i and j were passed to the hybrid, (2) inbred i came through but not inbred j, (3) inbred j came through but not inbred i, and (4) neither inbred came through. Assuming independence, these have respective probabilities P^2, $P(1 - P)$, $P(1 - P)$, $(1 - P)^2$. In the case $P = 0.50$, all of these probabilities equal 0.25.

An instance of the law of total probability (Sec. 5.3, Berry 1996) is that the probability of observing a hybrid's alleles is the average of the conditional probability of this event given the above four cases. The simplest of the four cases is the first possibility. Assuming the hybrid's alleles are passed down directly from both inbreds, the probability of observing the hybrid's genotype is either 1 or 0 depending on whether the hybrid shares both inbreds' alleles. (It is especially easy when both inbreds are homozygous.) The other three cases require an assumption regarding the possibility that an inbred's allele is not passed to the hybrid but is interrupted by a mutation, a laboratory error, or intervening breeding. We regard such an allele as being selected from all known alleles with probability $1/$(number of alleles), where the number of alleles is the total number of alleles known to exist at the locus in question. An alternative approach would be to use the allelic proportions that are present in the database (or in another database). However, the lines in the database may not be randomly selected from any population. For example, a line that has been highly used in breeding would have many derivative lines in the database, in which case the frequencies of its alleles will be artificially inflated. Assuming equal probabilities for the various alleles at a given locus is robust in the sense that it is not affected by adding and dropping lines from the database.

There are many cases to consider when computing the probability of observing a hybrid's alleles, depending on the zygosity of the hybrid and the inbreds, and allowing for the possibility of missing alleles or "extra alleles" in the assessment of the hybrid and inbred genotypes. These possibilities are too numerous to list. Instead we give three simple examples. All the examples have homozygous inbreds, the most common case. And each of the three hybrids has two alleles, again the most common case. We suppose that the measured alleles for three SSRs and a particular trio of hybrid and ancestor inbreds are as we have indicated in Table 1.

For SSR 1 there are three known alleles, one in addition to alleles a and b that are listed for the three lines (hybrid, inbred i, and inbred j) in Table 1. For SSR 2 and SSR 3 there are two known alleles in addition to those listed. The calculations in the right half of Table 1 will now be explained. Implicit in calculating $P(SSR|i, j)$ is the assumption—required in both the numerator and denominator of Bayes' rule—that inbreds i and j are ancestors of the hybrid. Consider SSR 1. In case 1 above, both ancestors' alleles (as measured by the laboratory process) are assumed to pass to the index hybrid, and so in this case the hybrid is necessarily ab. The probability of observing the actual hybrid's genotype is 1 for case 1, as shown in Table 1. In case 2, we assume that inbred i's allele passes to the hybrid but inbred j's does not. Indeed, the hybrid has an a allele. The probability of observing a b as the other allele is $1/$(number of alleles) $= 1/3$, as shown in Table 1. Case 3 is similar. In case 4, neither ancestor allele is passed to the hybrid: the probability of observing the hybrid's genotype (or any heterozygous genotype) is $2(1, 3)(1, 3) = 2, 9$. Since $P = 0.50$, the overall (unconditional) probability in the rightmost column (17, 36) is the simple average of the four cases, as indicated in Table 1.

For SSR 2 and SSR 3 the calculations are similar. For SSR 2 there is some evidence against pair (i, j) being ancestors,

810   D. V. Berry et al.

TABLE 1

Probability of observing a hybrid's alleles using three sample SSRs and four possible combinations (cases) of alleles passed, assuming that Inbreds i and j are ancestors of the hybrid

SSR	No. of alleles	Hybrid	Inbred i	Inbred j	Probability of observing the hybrid's genotype				Overall probability $P(SSR\|i, j)$
					Case 1 i, j	Case 2 i, not j	Case 3 not i, j	Case 4 not i, not j	
1	3	ab	aa	bb	1	1/3	1/3	2/9	17/36
2	5	bd	bb	cc	0	1/5	0	2/25	7/100
3	6	ab	cc	dd	0	0	0	2/36	2/144

SSR, simple sequence repeat marker profile.

but it is not conclusive. For SSR 3 there is even less evidence favoring pair (i, j). It would not take many SSRs with evidence similar to that for SSR 3 to essentially rule out this pair—provided that other pairs are not similarly inconsistent.

To find the overall $P(SSRs|i, j)$, multiply the individual $P(SSR|i, j)$ over the various SSRs. There are purely computational issues to address. Each $P(SSR|i, j)$ is a number between 0 and 1. When there are a great many SSRs, the product of these numbers will be vanishingly small. To lessen problems with computational underflow, for each SSR we multiply $P(SSR|u, v)$ by the same constant for each pair (u, v): the inverse of the largest possible such probability. For example, since 17/36 is the largest probability for a heterozygous hybrid at an SSR having three alleles (as is the case for SSR 1 in Table 1), we multiply all factors $P(SSR/1u, v)$ by 36/17. To eliminate remaining problems with underflow, we do calculations using logarithms (adding instead of multiplying) and take antilogs at the end.

The probability $P(SSR|u, v)$ is calculated for all (u, v) pairs and summed over all possible pairings in the database, including that for the inbred pair under consideration: (i, j). This gives the denominator in the expression for $P(i, j|SSRs)$.

To determine the probability that any particular inbred, say Inbred i, is the closest ancestor of the index hybrid, sum $P(SSR|i, v)$ over all inbreds v with $v = j$. Call this $P(i|SSRs)$. The maximum of $P(i|SSRs)$ for any inbred i is 1. But since there is one closest ancestor on each side of the family, the sum of $P(i|SSRs)$ over all inbreds i is 2. If there is a particular pair (i, j) for which $P(i, j|SSRs)$ is close to 1, then both $P(i|SSRs)$ and $P(j|SSRs)$ separately will be close to 1.

SSR data: DNA was extracted from 54 maize hybrids and from 586 maize inbreds. All of the hybrids and most inbreds are proprietary products of Pioneer Hi-Bred International; some important publicly bred inbred lines were also included. The inbred parents and grandparents of each hybrid were included within the set of inbreds. Other inbreds that were genotyped include many that are highly related by pedigree to parents and grandparents of the hybrids. The hybrids were chosen because each has a pedigree that is known to us and collectively they represent a broad array of diversity of maize germplasm that is currently grown in the United States ranging from early to late maturity.

A total of 195 SSR loci were used in this study following procedures described in Smith et al. (1997), but modified as described below. SSR loci were chosen on the basis that they individually have been shown to have a high power of discrimination among maize inbred lines and collectively they provide for a sampling of diversity for each chromosome arm. Of these SSR loci, the following numbers (in parentheses) were located on individual maize chromosomes as follows: 1 (35), 2 (26), 3 (22), 4 (20), 5 (16), 6 (9), 7 (6), 8 (18), 9 (12), and 10

(14): 17 SSR loci have not yet been mapped. The correlations among the loci are unknown and are irrelevant for our methodology.

Sequence data for primers that allow many of these (and other) SSR loci to be assayed are available at website http://www.agron.missouri.edu. All primers were designed to anneal and amplify under a single set of conditions for PCR in 10-μl reactions. Genomic DNA (10 ng) was amplified in 1.5 mM MgCl₂, 50 mM KCl, 10 mM Tris-Cl (pH 8.3) using 0.3 units AmpliTaq Gold DNA polymerase (PE Corporation) oligonucleotide primer pairs (one primer of each pair was fluorescently labeled) at 0.17 μM and 0.2 mM dNTPs. This mixture was incubated at 95° for 10 min (hot start): amplified using 45 cycles of denaturation at 95° for 30 sec, annealing at 60° for 30 sec, extension at 72° for 35 sec; and then terminated at 72° for 10 min. A water bath thermocycler manufactured at Pioneer Hi-Bred International was used for PCR reactions. PCR products were prepared for electrophoresis by diluting 3 μl of each product to a total of 27 μl using a combination of PCR products generated from other loci for that same maize genotype (multiplexing) and/or dH20. Dilution of 1.5 μl of this mixture to 5 μl with gel loading dye was performed: it was then electrophoresed at 1700 V for 1.5 hr on an ABI model 377 automated DNA sequencer equipped with GENE-SCAN software v. 3.0 (PE-Applied Biosystems, Foster City, CA).

PCR products were sized automatically using the "local Southern" sizing algorithm (ELDER and SOUTHERN 1987). After sizing of PCR products using GeneScan, alleles were assigned using Genotyper software (PE-Applied Biosystems). Generally, allele assignations for each locus were made on the basis of histogram plots consisting of 0.5-bp bins. Breaks between the histogram plots of >1 bp were generally considered to constitute separation between allele bins: however, other criteria, such as the presence of the nontemplate-directed addition of adenine (+A addition) and naturally occurring 1-bp alleles, were used on a marker-by-marker basis to define the allele dictionary. All allele scores were made without knowing the identities of the maize genotypes.

RESULTS

Table 2 presents the probability of closest ancestry of the top five ranking inbred lines for each of 5 hybrids at $P = 0.50$ (Table 2A) and $P = 0.99$ (Table 2B). Probabilities of ancestry are shown for all 54 hybrids and the top ranking inbreds in Figure 1: $P = 0.50$ (Figure 1a) and $P = 0.99$ (Figure 1b). Results for the hybrids presented in Table 2 are featured at the top of Figure 1.

Probability of Ancestry Using SSR   S17

TABLE 2

Probability of ancestry of five hybrids using data obtained from 30, 100, and 195 SSR loci

Hybd.	30 loci			100 loci			195 loci		
	Inbd.	Prob.	SE	Inbd.	Prob.	SE	Inbd.	Prob.	SE
A. Assuming P = 0.50									
3417	SP1	0.9607	0.0125	P1	0.8749	0.0252	P1	1.0000	E-07
	P2	0.3077	0.1963	P2	0.8141	0.2235	P2	0.9957	0.0033
	D2P2	0.1016	0.1038	D1P2	0.1839	0.2235	D1P2	0.0043	0.0033
	D1P2	0.0907	0.0927	SP1	0.1243	0.025	D2P2	E-06	E-06
	P1	0.032	0.0125	D1P1	0.0009	0.0002	SP1	E-06	E-07
3525	P1	0.8343	E-07	P1	0.9999	<E-20	P1	1.0000	<E-20
	P2	0.8186	E-07	P2	0.5437	<E-20	P2	0.9635	0.0528
	D1P2	0.1699	E-07	D1P2	0.4563	<E-20	D1P2	0.0365	0.0528
	CP1	0.1441	E-07	CP1	E-07	E-18	SP1	E-15	<E-20
	CP2	0.0110	E-08	SP1	E-07	<E-20	CP2	E-16	<E-20
3556	P1	1.0000	E-06	P1	0.9999	E-10	P1	1.0000	<E-20
	P2	0.9616	E-08	P2	0.9997	E-10	P2	1.0000	<E-20
	D1P2	0.0340	E-10	D1P2	0.0003	E-14	D1P2	E-09	<E-20
	CP2	0.0043	E-09	D2P2	E-05	E-15	D2P2	E-14	<E-20
	D2P2	0.0002	E-10	D3P2	E-06	E-17	GCP2	E-17	E-17
3905	D1P1	0.9822	E-08	D1P1	0.9803	0.0058	P1	1.0000	E-03
	SP2	0.4927	E-07	SP2	0.6230	0.0076	D1P2	1.0000	E-06
	D2P2	0.2836	E-07	D1P2	0.2821	0.0617	D2P2	E-06	E-06
	D1P2	0.1622	E-07	D2P2	0.1317	0.0372	P2	E-07	E-13
	P2	0.0565	E-07	P1	0.0197	0.0058	D3P2	E-10	E-16
3940	P2	0.9997	0.0001	P2	0.9999	E-05	P2	1.0000	E-09
	D1P2	0.9203	0.0009	P1	0.9970	0.0011	P1	1.0000	E-09
	P1	0.0648	E-05	D1P2	0.0030	0.0011	D1P2	E-11	E-11
	D1P1	0.0127	E-05	D2P2	0.0001	E-05	DP1P2	E-17	E-17
	DP1P2	0.0014	0.0009	DP1P2	0.0001	E-07	D2P2	E-19	E-18
B. Assuming P = 0.99									
3417	SP1	0.9995	0.0001	P1	0.9999	E-05	P1	0.9999	E-08
	P2	0.8836	0.1658	P2	0.9938	0.0107	P2	0.9999	E-08
	D1P2	0.0722	0.1029	D1P2	0.0061	0.0107	D1P2	E-11	E-11
	D2P2	0.0441	0.0628	D1P1	E-05	E-06	D2P2	E-14	E-14
	P1	0.0001	0.0001	SP1	E-05	0	SP1	E-20	E-21
3525	P1	0.9999	0	P1	0.9999	0	P1	1.0000	0
	P2	0.8991	0	D1P2	0.9749	0	P2	0.6135	0.4446
	D1P2	0.1008	E-11	P2	0.025	0	D1P2	0.3864	0.4446
	CP1	E-05	0	D2P2	E-20	0	CP2	E-48	0
	CP2	E-06	E-17	SP1	E-24	0	D2P2	E-49	0
3556	P1	1.0000	0	P1	1.0000	0	P1	0.9999	0
	P2	0.9996	0	P2	0.9999	0	P2	0.9999	0
	D1P2	0.0003	0	D1P2	E-09	0	D1P2	E-22	0
	D1P1	E-11	0	D3P1	E-21	0	D2P1	E-49	0
	D2P1	E-13	0	D2P1	E-21	0	D3P1	E-54	0
3905	D1P1	0.9999	0	D1P1	0.9999	E-08	P1	1.0000	E-09
	P2	0.9992	0	P2	0.9999	E-06	P2	0.9947	E-09
	SP2	0.0006	0	D1P2	E-06	E-06	D1P2	0.0052	E-11
	D1P2	E-05	0	SP2	E-07	E-13	D2P2	E-18	E-18
	D2P2	E-06	0	D2P2	E-09	E-10	D1P1	E-25	E-25
3940	P2	0.9999	E-08	P2	1.0000	E-08	P1	1.0000	E-09
	D1P2	0.9999	E-08	P1	0.9999	E-05	P2	1.0000	E-09
	P1	E-06	E-13	D1P2	E-05	E-05	D1P2	E-24	E-24
	D1P1	E-08	E-15	D2P2	E-12	E-11	DP1P2	E-14	E-14
	DP1P2	E-12	E-12	DP1P2	E-21	E-21	D2P2	E-50	E-49

Hybd., hybrid; Inbd., inbred; Prob., probability; SE, standard error, referring to the variability in the results of the runs: P1, parent one; P2, parent two; SP1 SP2, full sibling of parent one parent two; D1P1/D1P2, derivatives of parent one parent two, index i for distinct inbred lines; DP1P2, derivatives of both parent one and parent two.

a
Hybrids

Probability of ancestry (log₁₀)

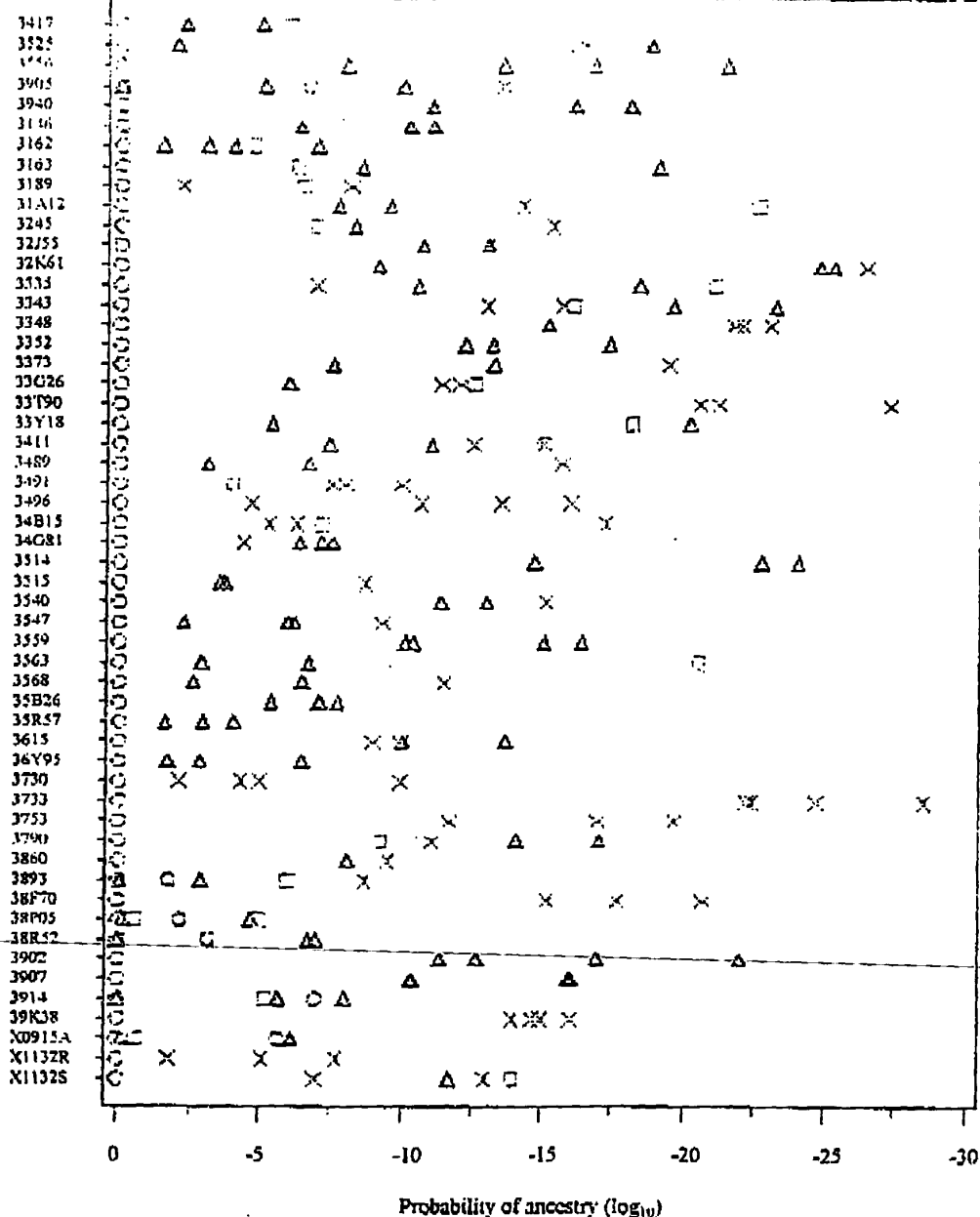△ △ △   D-P1/P2   ⌄ ✕ ⌐   Others   ⊏ ⌁ ⊐   Parent   ⌁ ⌁ ⊏   S-P1/P2

Figure 1.—(a) Probabilities of ancestry, assuming P = 0.30, for all 54 hybrids and top ranking inbreds—those with probability of ancestry at least 10⁻³. (b) Probabilities of ancestry, assuming P = 0.99, for all 54 hybrids and top ranking inbreds—those with probability of ancestry at least 10⁻³.
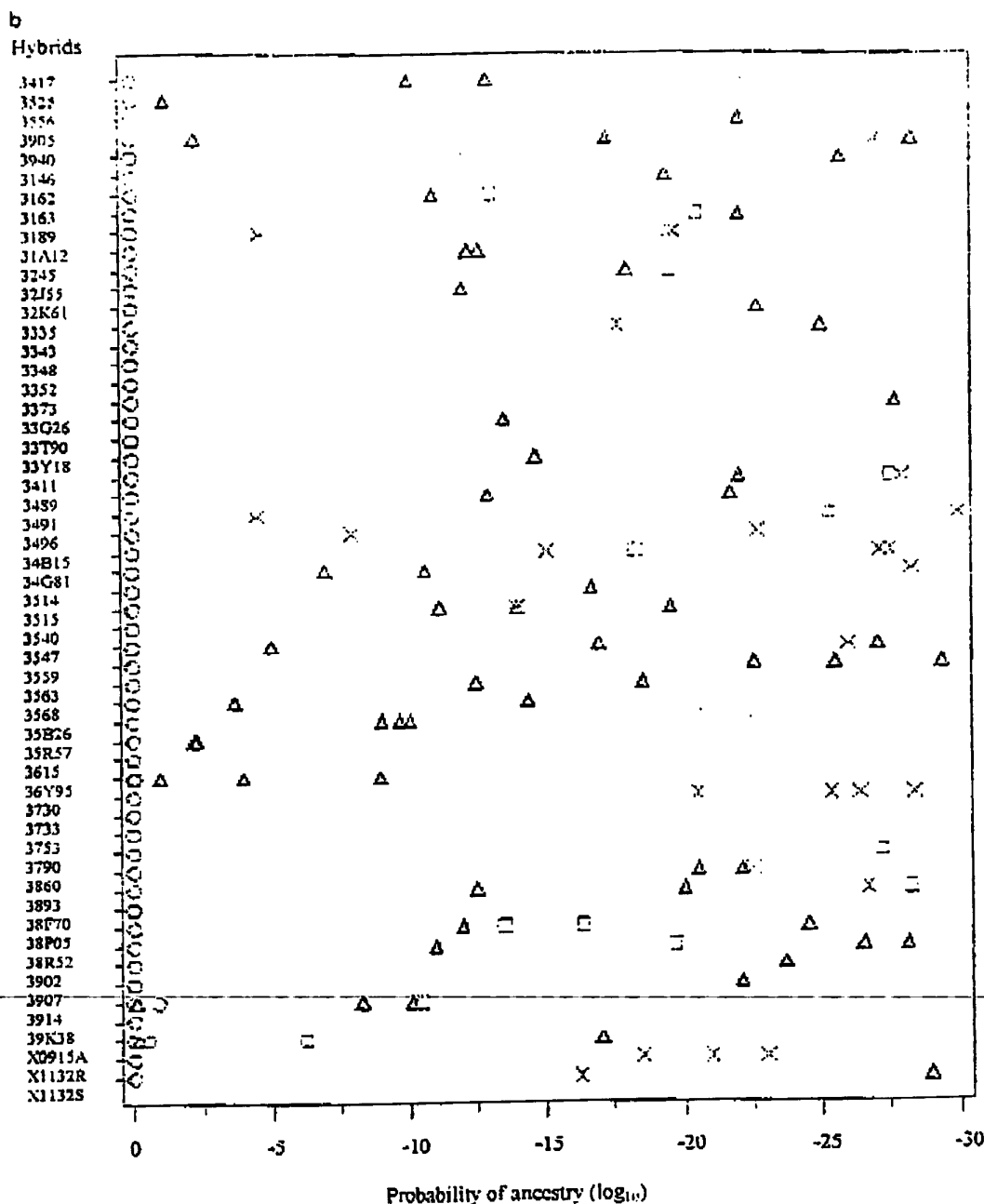
FIGURE 1.—Continued.

When the algorithm used $P = 0.50$, the two correct parents were identified as highest in probability for 48 (89%) hybrids (Figure 1). For each of 6 hybrids (3893, 38P05, 38R52, 3905, 3914, and X0915A), one parent ranked in the top two places. The other parent was supplanted either by a sister inbred or by an inbred that was a direct progeny of that parent. Overall, 102 (94%) of 108 parental inbreds were correctly identified. For hybrids where both parents ranked first or second, the range of probabilities for parental lines that ranked first from among all other inbreds ranged from 1.0000 to 0.9997; parental lines ranking second ranged from

820   D. A. Berry et al.

1,0000 to 0.9653. For 35 hybrids, both parents had probabilities of ancestry in excess of 0.999. Probabilities of ancestry for nonparents that ranked in first or second places were from 0.9999 to 0.7054. For the majority of hybrids, the probability of the third ranked hybrid se ranked nonparental inbred was at or below E-06. This indicates that there is usually very little uncertainty about closest ancestors.

When the algorithm used $P = 0.99$ to examine each of the 54 hybrids, both parents were correctly identified for 52 (96%) of hybrids and for 98% (102/104) of the parents across all hybrids (Figure 1). Two hybrids (3914 and X0915A), in which one parent was not ranked in the top two, were also in the subset not ranked in the top two assuming $P = 0.50$ (above). In both cases their ranks improved (both to third rank) and the actual parent was supplanted by an inbred that was a direct progeny of the corresponding parental line. For 49 hybrids, both parents had probabilities of ancestry in excess of 0.999. Among the 5 hybrids having a parent ranking second with a probability of ancestry below 0.999, the lowest of these probabilities was 0.8975 and the highest probability for a third ranking nonparent was 0.1023. For most hybrids the probability for the third and highest ranked nonparental inbred was at or below E-10.

Table 2 also addresses data analysis in circumstances where heterozygous loci occur in inbred lines or where a hybrid is scored for the presence of more than two alleles per locus. The presence of more than a single allele per locus in inbred lines is an infrequent occurrence in well-maintained inbred development and seed increase programs but is possible because ~3–5% of loci can still be segregating and unintended pollination from genotypes not designated as parents of the hybrid can occur. For hybrids, more than two alleles per locus can be scored when DNA is extracted from a bulk of individual plants and because inbred parents are not homozygous due either to residual heterozygosity or to contamination or because one or more direct parents of the hybrid are themselves hybrids. The presence of more than one allele per locus in an inbred line and more than two alleles per locus in a hybrid therefore can be accommodated by multiple runs of the algorithm, each with a random choice of two alleles per locus. Consequently, standard errors in the case of analyzing data from 195 loci tend to be very small because there were few loci where an inbred or hybrid sample (from a bulk of individual plants) was scored for more than two alleles.

Marshall, et al. (1998) have drawn attention to errors that can be encountered in genotyping surveys. These errors include missing data, null alleles, and typing errors. We therefore investigated the robustness of the algorithm by examining the effects of modifications in the data for five hybrids (3417, 3525, 3556, 3905, and

3940). First, we reduced the number of SSRs used, from the full set of 195 to 100 and then to 50 (Table 2). Use of 50 loci generated incorrect rankings of one parent for each of two hybrids (3417 and 3940) and for both parents of one hybrid 3905. All of these most highly ranked nonparental inbreds were closely related to the true parents for each of the respective hybrids: six different inbred lines were involved. Four were direct progeny of the true parents (one with additional backcrosses from the true parent) and two were full sisters (from a cross of highly related inbreds) of the actual parent of the hybrid. Using 100 loci resulted in correct parental rankings for all hybrids except for 3905 where neither parent ranked in first or second place. Four inbreds outranked the true parents of 3905. All four nonparents were closely related to the respective true parents; three were direct progeny of the true parent of the hybrid (one with additional backcrossing to that parent) and one was a full sister of the true parent. Use of data from all 195 loci corrected the placement for one of the parents of hybrid 3905. Two inbreds that were not parents of this hybrid remained ranked more highly than one of the true parents. Both were direct progeny of that parent, and one of these inbreds had additional backcrossing to that parent in its pedigree.

To address the consequences of laboratory and other sources of error, we artificially compromised data quality beyond the level originally provided by eliminating specific proportions of alleles that had been scored (establishing scenarios where various numbers of SSR alleles were not scored) and by misscoring other alleles (establishing scenarios where various numbers of SSR alleles were scored incorrectly). We also combined the scenarios of missing data and wrongly scored data. Table 3 contains a summary of the results of making these modifications in the data. For all modifications we used data from all SSR loci and we also randomly chose SSR loci to create subsets of 50 and 100 loci. In each case, the program was run 20 times for each hybrid/set of loci. When all 195 loci were examined, replications differed only according to the particular choice of alleles for loci where more than two alleles had been scored.

To evaluate robustness in the face of missing data or mistyped data, we simulated individual and combined categories of these data in the hybrid and all inbred lines at levels of 2, 5, 10, and 25% of the alleles for each of five hybrids and all inbreds beyond the level of error as originally scored by the laboratory. We examined the effects of these levels and types of error for three sizes of database: 50 loci, 100 loci, and all 195 scored loci. The same five hybrids considered in Table 2 were investigated: 3417, 3525, 3556, 3905, and 3940. One of these hybrids (3905) was chosen because one of its parents did not rank among the top two places even when the complete and unmodified data from all SSR loci were used.

Examples of robustness in the face of additional error

TABLE 3

Number of parents ranked in first and second positions (maximum is 2)

Type of simulated data	% level change	No. of loci:	Hybrid 3417 50	3417 100	3417 195	3525 50	3525 100	3525 195	3556 50	3556 100	3556 195	3905 50	3905 100	3905 195	3940 50	3940 100	3940 195	Mean % max.
Missing	0		1	2	2	2	2	2	2	2	2	0	0	1	1	2	2	77
	2		1	2	2	2	1	2	2	2	2	0	0	1	1	2	2	73
	5		1	2	2	2	2	2	2	2	2	0	0	1	1	2	2	77
	10		1	2	2	2	2	2	2	2	2	0	0	1	1	2	2	77
	25		0	2	2	1	1	1	2	2	2	0	0	1	1	1	1	57
	Mean % max.		40	100	100	90	80	90	100	100	100	0	0	50	50	90	90	
Misscored	0		1	2	2	2	2	2	2	2	2	0	0	1	1	2	2	77
	2		1	2	2	2	2	2	2	2	2	0	0	1	1	2	2	77
	5		1	2	2	2	1	2	2	2	2	0	0	1	1	2	2	73
	10		1	1	2	2	2	2	2	2	2	0	0	1	1	2	2	73
	25		1	0	2	1	1	2	1	2	2	0	1	1	1	2	2	63
	Mean % max.		50	70	100	90	80	100	90	100	100	0	10	50	50	100	100	
Missing plus misscored	0		1	2	2	2	2	2	2	2	2	0	0	1	1	2	2	77
	2		1	2	2	2	2	2	2	2	2	0	0	1	1	2	2	77
	5		1	2	2	2	1	1	2	2	2	0	0	1	1	2	2	70
	10		1	2	2	1	2	2	2	2	2	0	1	1	1	2	2	7
	25		0	1	2	0	0	0	0	0	0	0	0	1	0	0	1	17
	Mean % max.		40	90	100	70	70	70	80	80	80	0	10	50	40	80	90	
Overall mean			43	87	100	83	77	87	90	93	93	0	7	50	47	90	93	

Hybrids considered are the same as those in Table 2.

for five hybrids using subsets of 50 and 100 loci and all loci are shown in Table 3 where numbers of parents ranking into the top two places are presented. Degradation in the preferential ranking of parent inbreds at a level of 25% additional missing data was shown for one hybrid (3525) with usage of 50, 100, or all SSR loci. Degradation in the preferential ranking of parent inbreds at a level of 25% additional misscored data was shown for hybrid 3556. When both additional levels of missing and misscored data were simulated, degradation in the ability to preferentially rank inbred parents occurred for all hybrids and for all sets of SSR (50, 100, and 195 loci) except for hybrid 3417 when data from 195 SSR loci were used. Over all five hybrids, use of 100 loci improved robustness from the use of 50 loci; use of 195 loci further improved robustness for four hybrids (3417, 3525, 3905, and 3940). The degree of improvement was small, except for hybrid 3905.

We also ranked inbreds according to their probability of ancestry of hybrids when both parents and all inbred derivatives and full-sister inbreds of the respective inbred parents for each hybrid were excluded from the analysis. The results are too voluminous to present here but can be summarized as follows: Using $P = 0.50$, a grandparent of each respective hybrid ranked into first place for 41 (76%) hybrids; probabilities ranged from 0.4976 to 1.0 and most were above 0.9999. Other classes of inbreds that ranked in first position for probability of ancestry were inbreds derived directly by pedigree from a grandparent of the respective hybrid (DGP) for 13% of hybrids, inbreds derived directly by pedigree from a great-grandparent of the respective hybrid (DGGP) for 9% of hybrids, and one class (2% of hybrids) with an inbred ranked into first place that was directly related by pedigree to the great-great-grandparent of that hybrid. Inbreds that ranked in second position were related to the respective parents of the hybrid as follows: Thirty-one (57% of hybrids) were a grandparent of the respective hybrid, 11 (20%) were classed as DGP, 7 (13%) were DGGP, 1 (2%) was class DGGGP, and 4 (7%) were a great-grandparent (GGP) of the respective hybrid. Over all hybrids, two of the four grandparents ranked into first and second positions for 23 (43% of hybrids); three grandparents ranked into the first three positions for 5 (9% of hybrids). There were no instances where all four grandparents ranked into the first four positions. Thirty hybrids had a grandparent ranked into first position using $P = 0.99$. The number of grandparents ranked into the top five positions was 93 (compared to 108 when $P = 0.50$). The number of grandparents ranking into the top two positions was 55 (compared to 71 when $P = 0.50$). The mean probability of a grandparent that ranked into the first two positions was 0.9258 (SD = 0.1154) when $P = 0.50$ and 0.9980 (SD = 0.0104) when $P = 0.99$.

DISCUSSION

The prevalent use of paternity indices demonstrates that it is advantageous to have explicit probabilities of ancestry to distinguish among different pedigrees. Molecular marker profiles are rapidly becoming more extensive and cost effective to generate. Features that would advance the statistical analysis of molecular marker data to provide explicit probabilities of ancestry include the ability to calculate probabilities of ancestry where there is no *a priori* information as to the identity of one (usually the maternal) parent and robustness in the face of laboratory error.

Maize inbred lines and hybrids provide a very exacting set of materials for evaluating the discriminatory abilities of molecular data and statistical procedures that are employed to interpret those data. Hundreds of maize inbred lines of known pedigree together encompass a great diversity and complexity of pedigree relationships. Some inbred lines can be very highly related and genetically similar due to their derivation from common parentage including from parents that are themselves highly related. Consequently, relationship categories such as "sister" or "parent" when applied to maize inbreds usually refer to closer degrees of pedigree relationship and, thus, of germplasm and molecular marker profile similarity than those of the equivalently named classes of relationship for animal species. Most maize hybrids that are widely used in the United States today are constructed from pairs of inbred lines that are unrelated by pedigree, each inbred parent having been bred from a separate "pool" of germplasm. Various degrees of relatedness are possible between hybrids according to the pedigree relationships among their constituent inbred parents.

Using $P = 0.99$ in the algorithm is more specific for identifying parents than using $P = 0.50$. However, $P = 0.99$ is less robust for identifying other relatives, such as grandparents. When the algorithm was run at $P = 0.50$ there were 6 hybrids for which one parent did not rank among the top two most probable genotypes. For the remaining 48 hybrids the correct parents were identified even in circumstances where other candidate inbreds included not only full-sister lines bred from related parents but also inbreds even more closely related to the true parent by virtue of being backcross conversions of the inbred parent of the hybrid. For each of the 6 hybrids where a nonparent ranked above a true parent, that higher ranked inbred was always either a sister or progeny of the outranked true parent. The range of pedigree relationships as expressed by the Malécot coefficient of relatedness (MALÉCOT 1948) that was encompassed by pairs of true parents and more highly ranked inbred relatives of the true parents was from 0.8390 to 0.9680. A coefficient of 0.8390 approximates a relationship between inbred A and A' where

inbred A' has been bred from a cross of inbreds A and B with between one and two additional backcrosses of the parental inbred A. A Malécot coefficient of relationship of 0.9680 closely approximates a relationship between inbreds A and A″ where four additional backcrosses of parental inbred A follow the initial cross of inbreds A and B.

Running the algorithm at $P = 0.99$ in comparison to $P = 0.50$ raises the probability of ancestry for the parents while diminishing the probabilities for the third and lower ranking candidate inbred lines. Use of the algorithm at $P = 0.99$ increased both the percentage of hybrids with both parents ranked in the first two positions (from 89 to 96%) and the percentage of parental inbreds that were ranked first and second (from 94 to 98%). Two hybrids (3914 and X0915A) did not have both parents ranked first and second when the algorithm was run at $P = 0.99$. For both of these hybrids the nonparental inbred that outranked the true parent was itself a product by pedigree from the true parent that had been created by an additional four backcrosses of that parent; the Malécot coefficient of relationship between the parent of the hybrid and the inbred that outranked that parent for these two hybrids was 0.9686.

Robustness was tested by evaluating the effects of using data from different numbers of loci and by simulating additional levels of missing and misscored data up to combined levels of 25% error beyond that which was provided by the laboratory. From our experience, error rates of 5 to 10% can occur in SSR profiling of maize due chiefly to the combined effects of residual heterozygosity among seed lots and by deficiencies in the scoring of heterozygotes in hybrids. The additional levels of simulated error, therefore, include values (up to ~35% total error) that are well outside of our experience. For five hybrids that were examined, increasing the number of loci from 50 to 100 (with no additional missing or misscored data) did reduce the number of instances where inbreds that were not parents of a hybrid outranked the true parent from four to one. Nonetheless, all of these more highly ranked inbreds, although they were not themselves the true parents of the respective hybrid, were either direct progeny or full sisters of the true parent (Table 2). Consequently, if such degrees of error can be tolerated in respect of pedigrees for inbreds that are identified as parents of hybrids, then SSR data from 50 loci of equivalent discrimination ability are sufficient. Use of data from 50 loci also evidenced robustness in the face of up to 10% additional levels of either missing or misscored data: no degradation in the ability to identify a parent was apparent up to the level of 10% additional error except for 10% additional missing and misscored alleles for one hybrid (3525; Table 3). However, use of 100 loci increased the proportion of true parents that were correctly identified from 53% (for 50 loci) to 71% (mean correct parents over all

levels of error; Table 3). Use of data from 195 loci provided greater resiliency against additional levels of error. However, use of data from 195 loci was unable to provide resiliency against the negative effects of adding combined levels (at 25%) of both missing and misscored data (Table 3). At the 25% level of additional poor data integrity, inbreds that were not related to the true parent of the hybrid outranked the true parent for four of the five hybrids. Levels of missing or misscored data should, therefore, be kept below 15–20% (assuming a level of 5–10% error in the data we analyzed prior to simulating additional error).

We have previously examined the pedigrees of inbreds that are ranked into the first two positions when the true parents are removed from the list of candidate inbred lines. Usually, direct progeny or full sisters of the true parents then rank most highly (data not presented). We therefore examined the rankings of inbreds with respect to their ranking and probability of inclusion in the ancestry of each hybrid after the removal, not only of the true parents, but also of the progeny of the true parents and any full sisters of the true parents. In these circumstances the grandparents of the hybrids are ranked predominantly into top positions. Using $P = 0.50$, a grandparent ranked into first position for 76% hybrids and into second position for 57% hybrids; with $P = 0.99$ a grandparent ranked into first place in 56% of hybrids. At $P = 0.50$ two grandparents ranked into first and second positions for 43% hybrids and into the first three positions for an additional 9% hybrids. Most of the remaining inbreds that ranked into the top two positions were progeny of the grandparent. A total of 108 grandparents ranked into the top five positions when $P = 0.50$; 93 ranked into these positions when $P = 0.99$. Seventy-one grandparents ranked into the top two positions when $P = 0.50$; 55 grandparents ranked into these positions when $P = 0.99$. The mean probability of a grandparent in the top two positions was 0.9288 (SD 0.1454) when $P = 0.50$ and 0.9980 (SD 0.0104) when $P = 0.99$. Our algorithm was written to identify pairs of ancestors; alternative algorithms could be tailored to identify all grandparents once parents had been identified and removed from the list of candidate inbreds.

We have demonstrated the capability and robustness of an algorithm that can be used to show probability of parentage in circumstances where the *a priori* pedigree identity of neither parent is known. Exclusions are taken into account, thereby allowing parentage to be shown even when the two parents are not represented in the database of molecular profiles that are examined. Heterozygous candidate parents can be accommodated. The number of loci that is necessary to provide a reliable basis of determining pedigree is dependent upon the degree of relatedness among parents and nonparents and upon the discriminatory ability of the marker system

in the species of interest. Using $P = 0.99$ compared to $P = 0.50$ preferentially identified more true parents and with a greater difference of probability to third placed nonparents. If there is reasonable assurance that the parents are among the candidate list of inbreds, then $P = 0.99$ should be used; if greater robustness is required, then $P = 0.50$ should be used.

Applications of our algorithm include the identification of pedigrees among individuals of plant or animal species where molecular profile datasets exist that can be interpreted in terms of segregating alleles at individual marker loci and that provide a sufficient power of discrimination. Capabilities to generate large datasets of suitable molecular profile data are already available and are increasing rapidly with the advent of single nucleotide polymorphisms. One further application of our algorithm is to assist in the protection of intellectual property that is obtained on plant varieties or upon specific dams or sires of animals through the determination of pedigrees.

LITERATURE CITED

ALDERSON, G. W., H. L. GIBBS and S. G. SEALY. 1999 Parentage and kinship studies in an obligate brood parasitic bird, the brown-headed cowbird (*Molothrus ater*), using microsatellite DNA markers. J. Hered. 90: 182–190.

BEIN, G., B. DRILLER, M. SCHÜRMANN, P. M. SCHNEIDER and H. KIRCHNER, 1998 Pseudo-exclusion from paternity due to maternal uniparental disomy 16. Int. J. Leg. Med. 111: 328–330.

BERRY, D. A. 1991 Inferences using DNA profiling in forensic identification and paternity cases (with discussion). Stat. Sci 6: 175–205.

BERRY, D. A., 1996 *Statistics: A Bayesian Perspective.* Duxbury Press, Belmont, CA.

BERRY, D. A., and S. GEISSER. 1986 Inferences in cases of disputed paternity, pp. 353–382 in *Statistics and the Law,* edited by M. H. DeGroot, S. E. Fienberg and J. K. Kadane. Wiley Publishing, New York.

BOCKEL, B., P. NÜRNBERG and M. KRAWCZAK. 1992 Likelihoods of multilocus DNA fingerprints in extended families. Am. J. Hum. Genet. 51: 554–561.

BOWERS, J. E. and C. P. MEREDITH, 1997 The parentage of a classic wine grape. Cabernet Sauvignon. Nat. Genet. 16: 84–87.

CHAKRABORTY, R., T. R. MEAGHER and P. E. SMOUSE, 1988 Parentage analysis with genetic markers in natural populations. I. The expected proportion of offspring with unambiguous paternity. Genetics 118: 527–336.

CHAKRABORTY, R., L. JIN and Y. ZHONG. 1994 Paternity evaluation in cases lacking a mother and nondetectable alleles. Int. J. Leg. Med. 107: 127–131.

DEVLIN, B., K. ROEDER and N. C. ELLSTRAND, 1988 Fractional paternity assignment: theoretical development and comparison to other methods. Theor. Appl. Genet. 76: 369–380.

ELDER, J. K., and E. M. SOUTHERN, 1987 Computer-aided analysis of one dimensional restriction fragment gels, pp. 165–172 in *Nucleic Acid and Protein Sequence Analysis—A Practical Approach,* edited by M. J. Bishop and C. J. Rawlings. IRL Press, Oxford.

ELLSTRAND, N. C. 1984 Multiple paternity within the fruits of the wild radish. *Raphanus sativus.* Am. Nat. 123: 819–828.

GOTZ, K. and G. THALLER, 1998 Assignment of individuals to populations using microsatellites. J. Anim. Breed. Genet. 115: 53–61.

GINS, P. K., K. TROEMEL, P. STAPLETON and D. B. NIARKOWSKI. 1997 DNA analysis in disputed parentage: the occurrence of two apparently false exclusions of paternity, both at short tandem repeat (STR) loci, in the one child. Electrophoresis 18: 1656–1652.

HAMRICK, J. L. and A. SCHNABEL 1985 Understanding the genetic structure of plant populations: some old problems and a new approach. pp 50–70 in *Population Genetics in Forestry,* edited by H. R. Gregorius. Springer-Verlag, Heidelberg, Germany.

HELMINEN, P., V. JOHNSSON, C. EHNHOLM and L. PELTONEN. 1990 Proving paternity of children with deceased fathers. Hum. Genet. 87: 657–660.

LANG, J. W., R. K. AGGARWAL, K. C. MAJUMDAR and L. SINGH, 1998 Individualization and estimation of relatedness in crocodilians by DNA fingerprinting with a Bkm-derived probe. Mol. Gen. Genet. 238(1–2): 49–58.

MALÉCOT, G., 1948 *Les Mathématiques de l'Hérédité.* Masson et Cie. Paris.

MARSHALL, T. C., J. SLATE, L. E. B. KRUUK and J. M. PEMBERTON, 1998 Statistical confidence for likelihood-based paternity inference in natural populations. Mol. Ecol. 7: 639–655.

MEAGHER, T. R., 1986 Analysis of paternity within a natural population of *Chamaelirium luteum* (L). Identification of most-likely male parents. Am. Nat. 128: 199–215.

MEAGHER, T. R., and E. THOMPSON, 1986 The relationship between single parent and parent pair genetic likelihoods in genealogy reconstruction. Theor. Popul. Biol. 29: 87–106.

MILLER, P. S., 1975 Selective breeding programs for rare alleles: examples from the Przewalski's horse and California Condor pedigrees. Conserv. Biol. 9: 1262–1273.

PRIMMER, C. R., M. T. KOSKINEN and J. PIRONEN. 2000 The one that did not get away: individual assignment using microsatellite data detects a case of fishing competition fraud. Proc. R. Soc. Lond. B Biol. Sci. 267: 1699–1704.

RANNALA, B., and J. L. MOUNTAIN, 1997 Detecting immigration by using multilocus genotypes. Proc. Natl. Acad. Sci. USA 94: 9197–9201.

SEFC, K. M., H. STEINKELLNER, J. GLOSSL, S. KAMPFER and F. REGNER, 1998 Reconstruction of a grapevine pedigree by microsatellite analysis. Theor. Appl. Genet. 97: 227–231.

SMITH, J. S. C., E. C. L. CHIN, H. SHU, O. S. SMITH, S. J. WALL *et al.,* 1997 An evaluation of the utility of SSR loci as molecular markers in maize (*Zea mays* L.) comparisons with data from RFLPs and pedigree. Theor. Appl. Genet. 95: 163–173.

SMOUSE, P. E., and T. R. MEAGHER. 1994 Genetic analysis of male reproductive contributions in *Chamaelirium luteum* (L.) grav (Liliaceae). Genetics 136: 313–322.

TAYLOR, A. C., A. HORSUP, C. N. JOHNSON, P. SUNNUCKS and B. SHERWIN, 1997 Relatedness structure detected by microsatellite analysis and attempted pedigree reconstruction in an endangered marsupial, the northern hairy-nosed wombat *Lasiorhinus krefftii.* Mol. Ecol. 6: 9–19.

THOMPSON, E., and T. R. MEAGHER. 1987 Parental and sib likelihoods in genealogy reconstruction. Biometrics 43: 585–600.

VANKAN, D. M., and M. J. FADDY. 1999 Estimations of the efficacy and reliability of paternity assignments from DNA microsatellite analysis of multiple-sire matings. Anim. Genet. 30: 355–361.

WETTON, J. H., D. T. PARKIN and R. E. CARTER. 1992 The use of genetic markers for parentage analysis in *Passer domesticus* (house sparrows). Heredity 69: 243–254.

WHITE, E., J. HUNTER, C. DUBETZ, R. BROST, A. BRATTON *et al.*, 2000 Microsatellite markers for individual tree genotyping: application in forest crime prosecutions. J. Chem. Technol. Biotechnol. 75: 923–926.

Communicating editor: Z-B. ZENG

What is an "Essentially D rived Variety"?

The concept of essentially derived variety was introduced into the 1991 Act of the UPOV Convention in order to avoid plagiarism through mutation, multiple back-crossing and to fill the gap between Plant Breeder's Rights and patents, gap which was becoming important due to the development of the use of patented genetic traits in genetic engineering.

An essentially derived variety is a variety which is distinct and predominantly derived from a protected initial variety, while retaining the essential characteristics of that initial variety.

As indicated as an example in the UPOV Convention, essentially derived varieties may be obtained by the selection of a natural or induced mutant, or of a somaclonal variant, the selection of a variant individual from plants of the initial variety, back-crossing, or transformation by genetic engineering.

The commercialization of an essentially derived variety needs the authorization of the owner of the rights vested in the initial variety.

The concept of essentially derived variety does not at all abolish the Breeder's Exemption, as free access to protected plant varieties for breeding purposes is maintained. It is not a threat to biodiversity. On the contrary, it favors biodiversity, encouraging breeders developing and marketing original varieties.

Attachment C
Serial No. 09/759,800

http://www.worldseed.org/FAQedv.htm   11/22/2002

www. UPOV. ORG

INTERNATIONAL CONVENTI N

FOR THE

PROTECTION OF NEW VARIETIES OF PLANTS

of December 2, 1961, as revised

at Geneva on November 10, 1972,

on October 23, 1978, and

on March 19, 1991

adopted by the Diplomatic Conference

on March 19, 1991

reproduced from UPOV Publication No. 438(E)

issue No. 63 of "Plant Variety Protection"

1991 Act of the Convention

Article 12
Examination of the Application

Any decision to grant a breeder's right shall require an examination for compliance with the conditions under Articles 5 to 9. In the course of the examination, the authority may grow the variety or carry out other necessary tests, cause the growing of the variety or the carrying out of other necessary tests, or take into account the results of growing tests or other trials which have already been carried out. For the purposes of examination, the authority may require the breeder to furnish all the necessary information, documents or material.

Article 13
Provisional Protection

Each Contracting Party shall provide measures designed to safeguard the interests of the breeder during the period between the filing or the publication of the application for the grant of a breeder's right and the grant of that right. Such measures shall have the effect that the holder of a breeder's right shall at least be entitled to equitable remuneration from any person who, during the said period, has carried out acts which, once the right is granted, require the breeder's authorization as provided in Article 14. A Contracting Party may provide that the said measures shall only take effect in relation to persons whom the breeder has notified of the filing of the application.

CHAPTER V
THE RIGHTS OF THE BREEDER

Article 14
Scope of the Breeder's Right

(1) [Acts in respect of the propagating material] (a) Subject to Articles 15 and 16, the following acts in respect of the propagating material of the protected variety shall require the authorization of the breeder:

   (i)   production or reproduction (multiplication),

   (ii)   conditioning for the purpose of propagation,

   (iii) offering for sale,

   (iv)   selling or other marketing,

   (v)   exporting,

   (vi)   importing,

   (vii) stocking for any of the purposes mentioned in (i) to (vi), above.

(b) The breeder may make his authorization subject to conditions and limitations.

(2) [Acts in respect of the harvested material] Subject to Articles 15 and 16, the acts referred to in items (i) to (vii) of paragraph (1)(a) in respect of harvested material, including entire plants and parts of plants, obtained through the unauthorized use of propagating material of the protected variety shall require the authorization of the breeder, unless the breeder has had reasonable opportunity to exercise his right in relation to the said propagating material.

1991 Act of the Convention   Page 7

(3) [Acts in respect of certain products] Each Contracting Party may provide that, subject to Articles 15 and 16, the acts referred to in items (i) to (vii) of paragraph (1)(a) in respect of products made directly from harvested material of the protected variety falling within the provisions of paragraph (2) through the unauthorized use of the said harvested material shall require the authorization of the breeder, unless the breeder has had reasonable opportunity to exercise his right in relation to the said harvested material.

(4) [Possible additional acts] Each Contracting Party may provide that, subject to Articles 15 and 16, acts other than those referred to in items (i) to (vii) of paragraph (1)(a) shall also require the authorization of the breeder.

(5) [Essentially derived and certain other varieties] (a) The provisions of paragraphs (1) to (4) shall also apply in relation to

(i) varieties which are essentially derived from the protected variety, where the protected variety is not itself an essentially derived variety,

(ii) varieties which are not clearly distinguishable in accordance with Article 7 from the protected variety and

(iii) varieties whose production requires the repeated use of the protected variety.

(b) For the purposes of subparagraph (a)(i), a variety shall be deemed to be essentially derived from another variety ("the initial variety") when

(i) it is predominantly derived from the initial variety, or from a variety that is itself predominantly derived from the initial variety, while retaining the expression of the essential characteristics that result from the genotype or combination of genotypes of the initial variety,

(ii) it is clearly distinguishable from the initial variety and

(iii) except for the differences which result from the act of derivation, it conforms to the initial variety in the expression of the essential characteristics that result from the genotype or combination of genotypes of the initial variety.

(c) Essentially derived varieties may be obtained for example by the selection of a natural or induced mutant, or of a somaclonal variant, the selection of a variant individual from plants of the initial variety, backcrossing, or transformation by genetic engineering.

Article 15
Exceptions to the Breeder's Right

(1) [Compulsory exceptions] The breeder's right shall not extend to

(i) acts done privately and for non-commercial purposes,

(ii) acts done for experimental purposes and

(iii) acts done for the purpose of breeding other varieties, and, except where the provisions of Article 14(5) apply, acts referred to in Article 14(1) to (4) in respect of such other varieties.

(2) [Optional exception] Notwithstanding Article 14, each Contracting Party may, within reasonable limits and subject to the safeguarding of the legitimate interests of the breeder, restrict the breeder's right in relation to any variety in order to permit farmers to use for propagating purposes, on their own holdings, the product of the harvest which they have obtained by planting,

1991 Act of the Convention   Page 8

Attorney Docket No. 1368

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicants:	Louis Brian Chapko and Todd Elliott Piper	Date:	Dec. 10, 2002
Serial No.:	09/759,800	Group Art Unit:	1638
Filed:	January 12, 2001	Examiner:	Ashwin D. Mehta
For:	"INBRED MAIZE LINE PH7JB"		

Assistant Commissioner for Patents
Washington, D.C. 20231

RULE 132 DECLARATION
OF
DR. STEPHEN SMITH

Sir:

I, Stephen Smith, PhD., do hereby declare and say as follows:

1.   I am skilled in the art of the field of the invention. I have a Ph.D. in Biochemical Systematics and Taxonomy of Maize and its Wild Relatives from Birmingham University. I have a M.Sc. in the Conservation and Utilization of Plant Genetic Resources from Birmingham University. I have a Bachelor of Science degree in Plant Sciences from London University. Since 1977 I have been engaged in the development, study and application of molecular markers to genetics, measuring genetic diversity and tracking pedigrees. I commenced this work at North Carolina State University as a post-doctoral research fellow. I have continued my engagement in these studies during my employment by Pioneer Hi-Bred from 1980 until the present. These studies have resulted in numerous scientific articles that have appeared in peer reviewed scientific literature.

2.   I have read and understood the Office Action in the above case dated September 4, 2002. This declaration is in response to the Examiner's rejection under, 35 U.S.C. § 102(e) as anticipated by or, in the alternative, under 35 U.S.C. § 103(a) as obvious over Piper (U.S. Patent No. 6,188,001).

3.   I have conducted an analysis of SSR marker data for inbred PH7JB and the inbred cited as prior art, PH1W0. Out of a total of 147 SSR loci examined, which allowed a sampling of each chromosome, there are 21 markers that show differences between PH7JB and PH1W0. This represents a difference for 14% for the markers

Attachment D
Serial No. 09/759,800

09/759,800

tested. Of these 21 markers, 13 were greater than 50 cM in distance, or unlinked on the genetic map.

4.   Upon crossing PH7JB to any other maize line and selfing successive filial generations, one would within the realm of what is statistically possible, obtain a progeny inbred maize line that retains genetic contribution from PH7JB. Assuming that (i) the cited prior art is used as the maize line to which PH7JB is crossed, (ii) that the only difference between PH7JB and PH1W0 are these 21 markers, and (iii) that all markers within a 50 cM distance will segregate together, then the odds of obtaining a PH7JB progeny inbred that is the same as PH1W0 after one cycle of breeding, is 1 in 2^{13} or 1 in 8,192. Statistically it is extremely unlikely that a PH7JB progeny, after one cycle of breeding, would be the same as PH1W0.

5.   Further, the assumptions made above vastly overstate the likelihood of breeding PH7JB from PH1W0. For example, it is common practice in quantitative genetics to determine the relation of plants by differences in markers. In doing so, one extrapolates that a percentage difference in markers is indicative of a difference in the whole genome. To assume that the only differences between PH7JB and PH1W0 are for these 21 markers, when 21 markers constitute 14% of the 147 SSR loci examined, is a gross and unrealistic assumption. Further the current maize genetic map only has approximately sixty 50cM units, so by applying this limitation the maximum number of independently segregating loci one could obtain, using the most different maize lines that could ever be found, is sixty. These assumptions result in an over estimate of the odds of breeding PH7JB from PH1W0.

6.   Given the difference in molecular markers between PH7JB and PH1W0, it is my expert opinion that PH7JB and PH1W0 are very distinct inventions. It is also my expert opinion that, within the realm of what is statistically possible, any progeny of PH7JB developed through crossing PH7JB with another plant will be distinct from PH1W0. Given the facts and based on my education and scientific experience, I believe that the invention as claimed is not obvious nor anticipated by Piper (U.S. Patent No. 6,188,001).

7.   I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

2

09/759,800

Date: _Dec 11ᵗʰ 2002_ By: _Stephen Smith_

Stephen Smith

3

AUG - 5 2002

ATCC

10801 University Blvd • Manassas, VA 20110-2209 • Telephone: 703-365-2700 • FAX: 703-365-2745

The American Type Culture Collection (ATCC) has received your deposit of seeds in connection with the filing of an application for patent. The following information is provided to fulfill Patent Office requirements.

Name and Address of Depositor: Pioneer Hi-Bred International, Inc.
Attn: Kim M. Hagemann
7100 NW 62nd Avenue
PO Box 1000
Johnston, IA 50131-1000

DOCKETED

AUG 15 2002

Deposited on Behalf of: Pioneer Hi-Bred International, Inc.

Date of Receipt of Seeds by the ATCC: July 8, 2002

Scientific Description	Depositor's Reference	Patent Deposit Designation
Inbred corn (maize) seed (B1PNE11966-01 RP)	PH7JB - /96:8	PTA-4531

The ATCC understands that:

1.   The deposit of these seeds does not grant ATCC a license, either express or implied, to infringe the patent, and our release of these seeds to others does not grant them a license, either express or implied, to infringe the patent.

2.   If these seeds should die or be destroyed during the effective term of the patent, it shall be your responsibility to replace them with living seeds of the same type. It is also your responsibility to supply a sufficient quantity for distribution for the deposit term.

Prior to the issuance of a U.S. Patent, the ATCC agrees in consideration for a one-time service charge, not to distribute these seeds or any information relating thereto or to their deposit except as instructed by the depositor or relevant patent office. After a relevant patent issues, and we are instructed to release the seeds, they will be made available for distribution to the public without any restrictions.

The ATCC agrees to maintain the seeds for a period of 30 years from deposit date, or 5 years after the most recent request for a sample, whichever is longer.

We will inform you of requests for the seeds for 30 years from date of deposit.

The seeds were tested July 16, 2002 and were viable.

American Type Culture Collection

By _____
Marie Harris, Patent Specialist
ATCC Patent Depository

Date: July 30, 2002